REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

PROMPT INJECTION EM PETIÇÕES JUDICIAIS E ACESSO À JUSTIÇA: INTEGRIDADE DOCUMENTAL, BOA-FÉ PROCESSUAL E OS LIMITES DA RESPOSTA SANCIONATÓRIA À MANIPULAÇÃO DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL NO PROCESSO CIVIL BRASILEIRO

PROMPT INJECTION IN COURT FILINGS AND ACCESS TO JUSTICE: DOCUMENT INTEGRITY, PROCEDURAL GOOD FAITH AND THE LIMITS OF THE SANCTIONING RESPONSE TO THE MANIPULATION OF ARTIFICIAL INTELLIGENCE SYSTEMS IN BRAZILIAN CIVIL PROCEDURE

INYECCIÓN DE INSTRUCCIONES (*PROMPT INJECTION*) EN ESCRITOS JUDICIALES Y ACCESO A LA JUSTICIA: INTEGRIDAD DOCUMENTAL, BUENA FE PROCESAL Y LÍMITES DE LAS SANCIONES POR LA MANIPULACIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN EL PROCEDIMIENTO CIVIL BRASILEÑO

Danilo Augusto Kanthack Paccini
Mestre, Emeron, Brasil
E-mail: danilo@tjro.jus.br

Cristiano Gomes Mazzini
Doutorando, Emeron, Brasil
E-mail: cristianomazzini@gmail.com

Resumo

A incorporação de sistemas de inteligência artificial às rotinas de leitura, síntese e análise de peças processuais, disciplinada pelo marco regulatório do Conselho Nacional de Justiça, inscreve-se no movimento de ampliação do acesso à justiça, ao prometer celeridade e redução de assimetrias informacionais. Esse mesmo movimento, contudo, inaugura categoria inédita de risco: a inserção deliberada de instruções ocultas em documentos processuais, destinadas a manipular o comportamento de sistemas de inteligência artificial que venham a processá-los, fenômeno conhecido na literatura técnica como prompt injection. O artigo analisa, a partir de estudo de caso único, decisão saneadora proferida em ação de obrigação de fazer contra operadora de plano de saúde, na qual o juízo identificou, em fonte ilegível ao olhar humano, comandos direcionados a eventual inteligência artificial que viesse a processar a petição inicial. O estudo articula quatro frentes de análise: a caracterização técnica do fenômeno, a partir da literatura de segurança de modelos de linguagem; a subsunção da conduta às cláusulas gerais de boa-fé objetiva, cooperação

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

e paridade de armas, e à vedação à litigância de má-fé do Código de Processo Civil; a releitura do episódio à luz da teoria das ondas, sustentando que a manipulação de sistemas de inteligência artificial no processo constitui prática antiacesso em dupla dimensão, por corromper a paridade de armas e por comprometer o acesso a uma ordem jurídica justa, dimensões das quais decorre risco proibicionista aos ganhos coletivos da tecnologia; e a discussão dos limites da resposta sancionatória em face da habilitação governada, proposta como alternativa ao bloqueio universal. Conclui-se que a resposta jurisdicional sancionatória, embora dogmaticamente adequada, é estruturalmente reativa e dependente da detecção humana, e que a resposta institucional consistente não é o bloqueio universal das ferramentas, mas a habilitação governada, categoria proposta neste artigo e articulada em camadas técnica, educacional e sancionatória.
Palavras-chave: Prompt injection. Acesso à justiça. Boa-fé processual. Inteligência artificial no Judiciário. Litigância de má-fé.

Palavras-chave: Prompt injection. Acesso à justiça. Boa-fé processual. Inteligência artificial no Judiciário. Litigância de má-fé.

Abstract

The incorporation of artificial intelligence systems into the reading, summarization and analysis of court filings, governed by the regulatory framework of the Brazilian National Council of Justice, is part of the movement to expand access to justice, promising speed and reduction of informational asymmetries. This same movement, however, introduces an unprecedented category of risk: the deliberate insertion of hidden instructions in procedural documents, designed to manipulate the behaviour of artificial intelligence systems that may process them, a phenomenon known in the technical literature as prompt injection. This article analyses, through a single case study, an interlocutory decision rendered in a claim against a health insurance operator, in which the court identified, in typeface illegible to the human eye, commands directed at any artificial intelligence system that might process the initial pleading. The study articulates four analytical fronts: the technical characterisation of the phenomenon, based on the language model security literature; the classification of the conduct under the general clauses of objective good faith, cooperation and equality of arms, and the prohibition of bad faith litigation of the Brazilian Code of Civil Procedure; the reinterpretation of the episode in light of the waves of access to justice, arguing that the manipulation of artificial intelligence systems in litigation constitutes an anti-access practice in a twofold dimension, as it corrupts equality of arms and compromises access to a just legal order, dimensions from which a prohibitionist risk to the collective gains of the technology arises; and the discussion of the limits of the sanctioning response in light of governed enablement, proposed as an alternative to universal blocking. It concludes that the sanctioning judicial response, although dogmatically sound, is structurally reactive and dependent on human detection, and that the

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

consistent institutional response is not the universal blocking of the tools, but governed enablement, a category proposed in this article and articulated in technical, educational and sanctioning layers.

Keywords: Prompt injection. Access to justice. Procedural good faith. Artificial intelligence in the Judiciary. Bad faith litigation.

Resumen

La incorporación de sistemas de inteligencia artificial a las rutinas de lectura, síntesis y análisis de escritos procesales, disciplinada por el marco regulatorio del Consejo Nacional de Justicia, se inscribe en el movimiento de ampliación del acceso a la justicia, al prometer celeridad y reducción de asimetrías informacionales. Este mismo movimiento, sin embargo, inaugura una categoría inédita de riesgo: la inserción deliberada de instrucciones ocultas en documentos procesales, destinadas a manipular el comportamiento de sistemas de inteligencia artificial que lleguen a procesarlos, fenómeno conocido en la literatura técnica como prompt injection. El artículo analiza, a partir de un estudio de caso único, una decisión de saneamiento procesal dictada en una acción de obligación de hacer contra una operadora de seguro de salud, en la cual el juzgado identificó, en fuente ilegible al ojo humano, comandos dirigidos a una eventual inteligencia artificial que llegara a procesar la demanda inicial. El estudio articula cuatro frentes de análisis: la caracterización técnica del fenómeno, a partir de la literatura de seguridad de modelos de lenguaje; la subsunción de la conducta a las cláusulas generales de buena fe objetiva, cooperación y paridad de armas, y a la prohibición de litigación de mala fe del Código de Proceso Civil brasileño; la relectura del episodio a la luz de la teoría de las olas de acceso a la justicia, sosteniendo que la manipulación de sistemas de inteligencia artificial en el proceso constituye práctica antiacceso en doble dimensión, al corromper la paridad de armas y al comprometer el acceso a un orden jurídico justo, dimensiones de las cuales deriva un riesgo prohibicionista para las ganancias colectivas de la tecnología; y la discusión de los límites de la respuesta sancionatoria frente a la habilitación gobernada, propuesta como alternativa al bloqueo universal. Se concluye que la respuesta jurisdiccional sancionatoria, aunque dogmáticamente adecuada, es estructuralmente reactiva y dependiente de la detección humana, y que la respuesta institucional consistente no es el bloqueo universal de las herramientas, sino la habilitación gobernada, categoría propuesta en este artículo y articulada en capas técnica, educativa y sancionatoria.

Palabras clave: Prompt injection. Acceso a la justicia. Buena fe procesal. Inteligencia artificial en el Poder Judicial. Litigación de mala fe.

1. Introdução

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

O acesso à justiça consolidou-se, desde a segunda metade do século XX, como a mais básica exigência de um sistema jurídico que pretenda garantir, e não apenas proclamar, os direitos de todos (Cappelletti; Garth, 1988). No Brasil, essa exigência tem assento constitucional expresso na garantia de inafastabilidade da jurisdição (art. 5º, XXXV, da Constituição Federal) e na garantia da razoável duração do processo (art. 5º, LXXVIII), e vem sendo progressivamente ressignificada, na literatura e na prática institucional, como acesso a uma ordem jurídica justa, e não como mera admissão formal ao aparato judiciário (Watanabe, 2019).

É nesse ínterim que deve ser lida a incorporação de sistemas de inteligência artificial às rotinas do Poder Judiciário brasileiro. A Resolução CNJ n. 615, de 11 de março de 2025, que revogou e substituiu a Resolução CNJ n. 332, de 2020, estabeleceu normas para o desenvolvimento, a governança, a auditoria, o monitoramento e o uso responsável de soluções de inteligência artificial no Poder Judiciário, com o objetivo declarado de promover a inovação tecnológica e a eficiência dos serviços judiciários em benefício dos jurisdicionados (CNJ, 2025), tendo sido posteriormente alterada pela Resolução CNJ n. 674, de 25 de março de 2026 (CNJ, 2026). Tribunais brasileiros vêm empregando modelos de linguagem para apoiar a triagem de processos, a elaboração de minutas e a produção de resumos de autos volumosos, movimento que, sob a ótica do acesso à justiça, promete atacar precisamente os gargalos de tempo e de escala que historicamente frustraram a efetividade da prestação jurisdicional.

Esse movimento, contudo, desloca para o processo civil um risco até então tratado quase exclusivamente pela literatura de segurança da informação: a possibilidade de que um documento textual, aparentemente neutro, contenha instruções ocultas capazes de alterar o comportamento de um sistema de inteligência artificial que venha a processá-lo. A esse fenômeno a literatura técnica denomina prompt injection, distinguindo entre a modalidade direta, na qual o próprio usuário insere a instrução manipuladora no comando dado ao sistema (Perez; Ribeiro, 2022), e a modalidade indireta, na qual a instrução é

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

embutida em conteúdo de terceiros que o sistema é levado a ler ou processar, sem que o usuário tenha conhecimento da manipulação (Greshake et al., 2023).
O fenômeno já não é hipótese de laboratório. Em julho de 2025, investigação jornalística e análise acadêmica subsequente documentaram a presença de instruções ocultas, redigidas em fonte branca ou microscópica, em dezessete manuscritos inicialmente identificados na plataforma arXiv, número ampliado para dezoito em levantamento acadêmico subsequente, com comandos dirigidos a eventuais revisores automatizados, do tipo que determinava a emissão de parecer exclusivamente favorável (Sugiyama; Eguchi, 2025; Lin, 2025).

O problema adquire contornos qualitativamente distintos quando a peça manipuladora é uma petição submetida ao contraditório e à apreciação de um juízo estatal. No Brasil, a primeira sanção noticiada ocorreu na Justiça do Trabalho, em caso apreciado pela 3ª Vara do Trabalho de Parauapebas, no Pará, no qual os patronos incluíram na petição inicial, em fonte branca sobre fundo branco, comando destinado a induzir eventual inteligência artificial empregada pela contraparte a produzir contestação superficial (Prado; Morimoto, 2026). O caso analisado neste artigo representa um segundo estágio, mais grave, do mesmo fenômeno. Decisão saneadora proferida por vara cível de comarca da Região Norte do país, em ação de obrigação de fazer cumulada com pedido de indenização por danos morais, ajuizada por segurado em face de operadora de plano de saúde para obtenção de cobertura de cirurgias reparadoras pós-bariátricas, identificou, ao final da petição inicial, após a assinatura dos patronos, em fonte de tamanho reduzido e cor branca sobre fundo branco, um conjunto extenso de comandos redigidos em três formatos de marcação estruturada, todos direcionados a orientar eventual sistema de inteligência artificial que viesse a analisar a peça a classificar os procedimentos pleiteados como reparadores e não estéticos, a aplicar entendimento jurisprudencial favorável à parte autora e a reforçar a urgência médica do caso, agora com mira não na inteligência artificial da contraparte, mas na que pudesse auxiliar o próprio órgão julgador. O prolator registrou não haver notícia de exemplar anterior tão complexo e extenso de prompt injection em peça processual submetida ao Poder Judiciário.

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

Diante desse quadro, formula-se o seguinte problema de pesquisa: como se caracteriza tecnicamente o fenômeno do prompt injection à luz da literatura de segurança de modelos de linguagem, em que medida a inserção de instruções ocultas direcionadas a sistemas de inteligência artificial em petição judicial configura, à luz do Código de Processo Civil brasileiro, ato atentatório à dignidade da justiça e litigância de má-fé, que efeitos essa prática produz sobre o acesso à justiça em suas dimensões individual e coletiva, e quais são os limites da resposta estritamente sancionatória diante de um risco cuja origem e cuja neutralização são, antes de jurídicas, técnicas?

O objetivo geral do artigo é analisar a fundamentação jurídica empregada pelo juízo para enquadrar a conduta nas categorias processuais pertinentes e discutir criticamente a suficiência dessa resposta frente à natureza do risco técnico subjacente e frente aos compromissos do sistema de justiça com a ampliação do acesso. Os objetivos específicos são quatro: caracterizar tecnicamente o fenômeno do prompt injection a partir da literatura de segurança de modelos de linguagem; situar a conduta identificada na decisão dentro da dogmática processual civil sobre boa-fé objetiva, lealdade, cooperação e deveres das partes; reinterpretar o episódio à luz da teoria das ondas de acesso à justiça, demonstrando sua nocividade em dupla dimensão e o risco sistêmico dela decorrente; e discutir por que a resposta institucional adequada ao fenômeno é a habilitação governada das ferramentas de inteligência artificial, e não o seu bloqueio universal.

O artigo está estruturado em cinco seções, além desta introdução. A segunda seção apresenta o referencial teórico, organizado em cinco subseções que tratam, respectivamente, da teoria do acesso à justiça e de sua dimensão tecnológica, dos fundamentos técnicos do prompt injection, da boa-fé processual objetiva e dos deveres de lealdade no direito brasileiro, da literatura sobre devido processo tecnológico, e do marco normativo institucional da inteligência artificial no Judiciário brasileiro. As duas últimas subseções compõem o pano de fundo normativo e teórico que sustenta os quatro objetivos específicos enunciados a seguir, e não constituem, isoladamente, objeto autônomo de investigação. A

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

terceira seção descreve a metodologia. A quarta seção apresenta a análise do caso, despersonalizada quanto às partes, incluindo a leitura do episódio como prática antiacesso. A quinta seção desenvolve o argumento central sobre os limites da resposta sancionatória e a inadequação do bloqueio universal como alternativa. A sexta seção conclui o trabalho.

2. Revisão da Literatura

O referencial teórico articula cinco eixos complementares. Os dois primeiros estabelecem as bases conceituais do problema, a teoria do acesso à justiça em sua dimensão tecnológica e a caracterização técnica do prompt injection. Os dois seguintes situam a conduta na dogmática processual civil brasileira e na literatura internacional sobre devido processo tecnológico. O quinto eixo descreve o marco normativo institucional que emoldura, no plano interno, o uso de inteligência artificial pelo Judiciário brasileiro.

2.1 Acesso à justiça: das ondas renovatórias à dimensão tecnológica

O Projeto de Florença, sintetizado na obra seminal de Cappelletti e Garth (1988), descreveu o movimento mundial de acesso à justiça por meio da metáfora das ondas renovatórias. A primeira onda voltou-se à superação dos obstáculos econômicos, por meio da assistência judiciária aos pobres. A segunda onda dirigiu-se à representação dos interesses difusos e coletivos, até então órfãos de tutela adequada. A terceira onda, mais ampla, propôs um novo enfoque de acesso à justiça, abrangendo a reforma dos procedimentos, a criação de mecanismos alternativos de solução de conflitos e a adaptação do processo ao tipo de litígio, sob a premissa de que as técnicas processuais servem a funções sociais e não constituem fim em si mesmas.

A literatura posterior propôs desdobramentos dessa tipologia. Economides (1999) sugeriu uma quarta onda, centrada na dimensão ética e na formação dos profissionais do direito, deslocando o foco do acesso dos cidadãos à justiça para o acesso dos próprios operadores jurídicos a uma concepção de justiça, e advertindo que de pouco vale ampliar portas de entrada se os responsáveis por operar o sistema não estiverem ética e tecnicamente preparados para fazê-lo. No

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

Brasil, Watanabe (2019) consolidou a compreensão do acesso à justiça como acesso à ordem jurídica justa, deslocando o eixo da admissão formal ao aparato judiciário para a qualidade e a justiça do resultado entregue ao jurisdicionado, enquanto Santos (2007) sublinhou que a demanda por justiça reprimida pela morosidade e pelo custo constitui forma de negação estrutural de direitos, exigindo uma transformação profunda da administração da justiça.

Mais recentemente, parte da literatura passou a identificar na tecnologia uma nova fronteira do movimento de acesso. Katsh e Rabinovich-Einy (2017) sustentam que a expansão das disputas nascidas em ambiente digital exige mecanismos de prevenção e resolução igualmente digitais, sob pena de se criar um contingente massivo de conflitos sem foro adequado. Susskind (2019) argumenta que a tecnologia, incluídos os sistemas de inteligência artificial, constitui a via mais promissora para reduzir a lacuna global de justiça, permitindo que sistemas judiciais processem, com qualidade e em tempo razoável, volumes de demandas que o modelo artesanal de adjudicação jamais absorveria. Nessa perspectiva, a adoção de inteligência artificial pelo Poder Judiciário não é um luxo tecnológico, mas um instrumento potencial de efetivação da garantia constitucional da razoável duração do processo e, portanto, um capítulo contemporâneo do próprio movimento de acesso à justiça.

Essa filiação não deve, contudo, ser confundida com adesão acrítica ao otimismo tecnológico. Parcela relevante da literatura adverte que sistemas automatizados podem reproduzir e amplificar vieses, operar sob opacidade refratária ao contraditório e aprofundar assimetrias entre litigantes habituais e eventuais, riscos documentados tanto na literatura internacional (Citron, 2008; Pasquale, 2015) quanto na produção processual brasileira dedicada à virada tecnológica (Nunes; Marques, 2018; Nunes; Lucon; Wolkart, 2022). A posição adotada neste artigo é, por isso, a de um otimismo condicionado: a tecnologia amplia o acesso à justiça se, e somente se, cercada de camadas de governança que assegurem transparência, supervisão humana e integridade dos insumos documentais submetidos ao processamento automatizado. É precisamente essa premissa condicional que torna inaceitáveis, a um só tempo, a manipulação dos

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

sistemas, que corrompe a condição de integridade, e o bloqueio universal, que sacrifica o ganho de acesso sem atacar a origem do risco, argumento que será desenvolvido na seção 5.

Essa filiação teórica, contudo, tem uma implicação de mão dupla que constitui uma das teses centrais deste artigo. Se a inteligência artificial integra a infraestrutura contemporânea do acesso à justiça, então a manipulação deliberada desses sistemas não é apenas um ilícito processual entre outros. É uma prática que ataca o próprio canal pelo qual o acesso se realiza, e que, por isso, deve ser lida também na chave da teoria do acesso, e não apenas na chave da dogmática das sanções processuais. Essa leitura será desenvolvida na seção 4.

2.2 Prompt injection como categoria técnica de risco em sistemas de linguagem

Perez e Ribeiro (2022) sistematizaram, no campo da segurança de modelos de linguagem, as técnicas de ataque por injeção de comandos, demonstrando empiricamente, por meio do arcabouço PromptInject, que agentes de baixa sofisticação técnica, porém suficientemente mal-intencionados, são capazes de desalinhar modelos de linguagem de grande porte mediante instruções manualmente elaboradas. Os autores identificaram duas modalidades básicas de ataque: o sequestro de objetivo, no qual a instrução maliciosa desvia o sistema da tarefa originalmente atribuída, e o vazamento de prompt, no qual o sistema é induzido a revelar suas instruções internas.

Greshake et al. (2023), em estudo amplamente citado na literatura subsequente, ampliaram o conceito ao demonstrar a viabilidade de ataques indiretos, nos quais o conteúdo manipulador não é fornecido pelo usuário que interage diretamente com o sistema, mas embutido em documentos, páginas web ou outros dados que o sistema é levado a recuperar e processar no curso de sua tarefa. Os autores descrevem esse vetor como especialmente perigoso porque explora a dificuldade estrutural desses sistemas em distinguir, dentro do texto que recebem, o que é instrução do que é dado a ser meramente descrito ou resumido. A organização setorial de segurança da informação OWASP (2023),

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

em sua sistematização de vulnerabilidades típicas de aplicações baseadas em modelos de linguagem, classifica o prompt injection como o risco de maior prioridade entre os catalogados.

A materialização do risco fora do laboratório está documentada. Em julho de 2025, reportagem da Nikkei Asia identificou dezessete manuscritos depositados na plataforma de preprints arXiv, de autores vinculados a quatorze instituições acadêmicas de oito países, contendo instruções ocultas dirigidas a eventuais revisores automatizados, redigidas em fonte branca sobre fundo branco ou em corpo microscópico, com comandos como o de emitir exclusivamente parecer favorável e não destacar aspectos negativos (Sugiyama; Eguchi, 2025). Análise acadêmica subsequente confirmou e ampliou o levantamento, identificando dezoito manuscritos e qualificando a prática como modalidade nova de má conduta científica, além de apontar que a vulnerabilidade se estende a qualquer sistema automatizado de processamento de textos, incluindo ferramentas de detecção de plágio e de indexação de citações (Lin, 2025). O episódio demonstra que a técnica de camuflagem identificada no caso judicial analisado neste artigo, fonte branca em corpo reduzido, não é invenção isolada, mas padrão já disseminado internacionalmente em contextos nos quais se presume o uso de inteligência artificial na avaliação de documentos.

A taxonomia de Greshake et al. (2023) é diretamente aplicável ao caso analisado neste artigo. O usuário do sistema de inteligência artificial, na hipótese, não seria o autor da manipulação, mas um terceiro, o juízo, a parte contrária ou qualquer operador que viesse a processar a petição com apoio de inteligência artificial, que teria seu sistema induzido a executar instruções camufladas como parte do documento a ser lido. Trata-se, portanto, de modalidade indireta, agravada pela circunstância de o vetor de ataque ser uma peça processual, protegida pela presunção de boa-fé e submetida a um procedimento estatal de resolução de conflitos.

2.3 Boa-fé objetiva, cooperação e paridade de armas no processo civil

A doutrina processual civil brasileira consolidou, após o Código de Processo Civil de 2015, a leitura da boa-fé processual como cláusula geral de natureza

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

objetiva, distinta da boa-fé subjetiva ligada ao estado mental do agente. Didier Jr. (2015) explica que o princípio da boa-fé processual, extraído do art. 5º do Código, constitui fonte normativa autônoma de deveres de conduta impostos a todos os sujeitos do processo, e não apenas às partes, cuja violação caracteriza abuso de direito processual, categoria dogmática distinta da mera violação de regra específica. Ao lado da boa-fé, o art. 6º positivou o dever de cooperação entre todos os sujeitos processuais para que se obtenha, em tempo razoável, decisão de mérito justa e efetiva, e o art. 7º assegurou às partes a paridade de tratamento em relação ao exercício de direitos e faculdades processuais, incumbindo ao juiz zelar pelo efetivo contraditório.

A articulação entre esses três dispositivos é relevante para o objeto deste estudo. A inserção de comandos ocultos dirigidos a sistemas de inteligência artificial viola simultaneamente os três planos: frustra a boa-fé objetiva, porque instaura uma comunicação processual dúplice, uma dirigida ao juízo humano e outra, oculta, dirigida à máquina que eventualmente o auxilie; nega a cooperação, porque instrumentaliza a infraestrutura de apoio à decisão em benefício exclusivo de uma das partes; e rompe a paridade de armas, porque confere a quem domina a técnica uma vantagem invisível e, por definição, insuscetível de contraditório, dado que a parte contrária não pode se contrapor a um argumento que não consegue ver.

Scarpinella Bueno (2015), ao comentar o art. 77 do Código de Processo Civil, situa o inciso VI do dispositivo, que veda a prática de inovação ilegal no estado de fato de bem ou direito litigioso, como cláusula de fechamento destinada a alcançar condutas não previstas taxativamente nos demais incisos do artigo, mas igualmente aptas a comprometer as condições de apreciação do direito em disputa. O comentário do autor à distribuição de responsabilidade entre parte e procurador, prevista nos §§ 6º e 8º do mesmo artigo, é diretamente relevante para a análise da decisão, que reconheceu a possível prática do ato atentatório de forma restrita aos procuradores signatários da peça, e não à parte autora enquanto tal.

Neves (2016) examina a hipótese de litigância de má-fé por conduta

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

temerária, prevista no inciso V do art. 80, destacando o caráter aberto da expressão, que autoriza sua aplicação a condutas processuais novas, ainda não tipificadas pelo legislador no momento da edição da norma. Essa leitura sustenta a subsunção, adotada na decisão analisada, de uma conduta tecnicamente inédita, a inserção de comandos direcionados a inteligência artificial, à cláusula geral de litigância de má-fé, sem necessidade de tipificação processual específica.

2.4 Devido processo tecnológico e responsabilização por uso indevido de sistemas automatizados

A literatura internacional sobre direito e tecnologia oferece moldura teórica complementar à dogmática processual civil brasileira. Citron (2008) cunhou a expressão devido processo tecnológico para designar o conjunto de garantias que devem acompanhar a automação de decisões e análises antes conduzidas exclusivamente por seres humanos, sustentando que a opacidade e a vulnerabilidade técnica de sistemas automatizados exigem mecanismos de transparência e de responsabilização equivalentes aos já consolidados para a atuação humana. Citron e Pasquale (2014) aprofundam essa análise ao discutir os riscos de sistemas de pontuação e classificação automatizada quando manipulados por atores interessados em seu resultado, hipótese estruturalmente análoga à manipulação de um sistema de inteligência artificial destinado a auxiliar a leitura de uma peça processual. Pasquale (2015) situa esses riscos no contexto mais amplo da opacidade algorítmica, argumentando que a confiança social em sistemas automatizados de tomada de decisão depende da possibilidade de auditar e contestar seu funcionamento. No Brasil, essa agenda foi recepcionada e desenvolvida pela literatura da virada tecnológica no direito processual, que examina os impactos da automação sobre as garantias processuais e adverte, em particular, para os riscos de atribuição de função decisória a sistemas enviesados e não auditáveis (Nunes; Marques, 2018; Nunes; Lucon; Wolkart, 2022), advertência que reforça, no plano nacional, a exigência de integridade dos insumos documentais submetidos a processamento automatizado.

Para os fins deste artigo, a contribuição decisiva dessa literatura é a

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

demonstração de que a integridade dos sistemas automatizados é condição de legitimidade do seu uso em contextos decisórios públicos. Se o devido processo tecnológico exige auditabilidade e contestabilidade, um cenário no qual peças processuais podem carregar instruções invisíveis e não detectadas compromete não apenas o caso concreto, mas a própria justificação institucional do emprego de inteligência artificial na jurisdição.

2.5 Governança institucional da inteligência artificial no Judiciário brasileiro

No plano normativo interno, o marco de referência é a Resolução CNJ n. 615, de 11 de março de 2025, que revogou a Resolução CNJ n. 332, de 2020, e estabeleceu normas para o desenvolvimento, a governança, a auditoria, o monitoramento e o uso responsável de soluções de inteligência artificial no âmbito do Poder Judiciário, estruturadas em torno de fundamentos como a supervisão humana efetiva, a proteção de dados pessoais, a formulação de soluções seguras com identificação e mitigação de riscos sistêmicos e a garantia de integridade e rastreabilidade dos dados utilizados (CNJ, 2025). A norma dialoga com a agenda doutrinária brasileira que, desde antes de sua edição, reivindicava parâmetros de governança e de contenção de riscos para o emprego de inteligência artificial no processo (Nunes; Lucon; Wolkart, 2022). A norma foi alterada pela Resolução CNJ n. 674, de 25 de março de 2026, que modificou a composição do Comitê Nacional de Inteligência Artificial do Judiciário, reforçando a integração entre a governança de inteligência artificial e a proteção de dados pessoais (CNJ, 2026).

Essa integração normativa entre governança algorítmica e proteção de dados pessoais ganha relevância direta no fenômeno analisado neste artigo, na medida em que o dado de saúde tratado pelo sistema de inteligência artificial que processa a petição judicial constitui categoria de dado pessoal sensível, sujeita a regime de tratamento mais restritivo. A camada técnica de triagem documental discutida na seção 5, ao impedir que instruções ocultas direcionem indevidamente o processamento automatizado de tais dados, cumpre função de salvaguarda que transcende a integridade processual, alcançando também a proteção de dados pessoais sensíveis.

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

No plano local, a decisão analisada apoiou-se na Nota Técnica n. 2/2025, editada pela Coordenadoria de Governança de Inteligência Artificial do tribunal de origem, cuja definição do fenômeno merece transcrição, conforme reproduzida no corpo da própria decisão:

> Prompt injection consiste na inserção deliberada de instruções ocultas ou manipulativas em documentos processuais, com o objetivo de alterar indevidamente o comportamento de sistemas de inteligência artificial encarregados da análise, classificação ou sumarização desses documentos, comprometendo a integridade e a confiabilidade dos resultados automatizados. Técnicas reconhecidas incluem: I - texto em fonte branca sobre fundo branco; II - instruções em metadados de arquivos PDF; III - comandos incorporados em camadas ocultas de documentos; IV - uso de fontes de tamanho microscópico ou ilegível; V - instruções em idiomas não usuais ou codificados (Nota Técnica n. 2/2025-CGIA, apud decisão analisada).

A existência dessa nota indica movimento institucional de resposta ao fenômeno e demonstra que a taxonomia técnica descrita na subseção 2.2 já foi internalizada pela governança de ao menos um tribunal brasileiro. Ela também evidencia, contudo, que a resposta disponível, até o momento, é predominantemente conceitual e reativa, dependente de detecção humana, e não uma camada técnica de triagem automática incorporada aos sistemas processuais eletrônicos, lacuna que será retomada nas seções 4 e 5.

3. Metodologia

A pesquisa filia-se ao paradigma interpretativista, por compreender o fenômeno jurídico como realidade socialmente construída, cuja inteligibilidade depende da atribuição de sentido às práticas dos atores processuais e institucionais, e adota percurso de raciocínio predominantemente abdutivo, na acepção originalmente formulada por Peirce (1934) e sistematizada na literatura de

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

metodologia de pesquisa como estratégia voltada à formulação da explicação mais plausível para um fato surpreendente (Saunders; Lewis; Thornhill, 2011): parte-se de um fato surpreendente, a identificação de comandos ocultos dirigidos a sistemas de inteligência artificial em petição judicial, para a formulação da hipótese explicativa que melhor o acomoda, qual seja, a de que a manipulação desses sistemas constitui prática antiacesso a exigir resposta institucional de governança, e não meramente sancionatória.

Dessa filiação decorrem a natureza qualitativa e exploratória do estudo e a consciência de seus limites: as conclusões não aspiram a generalização estatística, mas a generalização analítica, própria dos estudos de caso, na medida em que buscam gerar compreensão aprofundada de um fenômeno específico, e não inferência probabilística sobre uma população (Alves-Mazzotti, 2006).

A pesquisa é adequada ao estudo de fenômeno jurídico emergente para o qual ainda não existe literatura doutrinária consolidada. Adota-se o delineamento de estudo de caso único, uma das cinco abordagens de investigação qualitativa sistematizadas por Creswell (2014), justificado pelo caráter revelador do caso, hipótese na qual o caso singular franqueia ao pesquisador acesso a fenômeno até então inobservado e dotado, por isso, de valor analítico próprio (Flyvbjerg, 2006).

O caso ora analisado não constitui o primeiro registro brasileiro do fenômeno, pois há notícia de sanção anterior na Justiça do Trabalho, no caso apreciado pela 3ª Vara do Trabalho de Parauapebas, no Pará (Prado; Morimoto, 2026). Este estudo distingue-se do precedente, contudo, em três aspectos que justificam sua seleção: é, até onde foi possível apurar, o primeiro na jurisdição civil comum; é o mais complexo e extenso já documentado, com comandos redundantes em três formatos de marcação estruturada, circunstância registrada pelo próprio prolator; e tem como alvo não a inteligência artificial da contraparte, como em Parauapebas, mas a que pudesse auxiliar o próprio órgão julgador, o que desloca o ataque da esfera da lealdade entre as partes para a esfera da integridade da jurisdição.

A alegação de ineditismo na jurisdição civil comum foi submetida a controle metodológico mediante busca sistemática nas bases públicas de jurisprudência dos

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

tribunais brasileiros, com emprego dos descritores prompt injection, instrução oculta, comando oculto e inteligência artificial em associação com litigância de má-fé. Registra-se que a eventual busca negativa não exclui a existência de casos não indexados ou não noticiados, razão pela qual a afirmação de ineditismo é formulada em termos de ausência de registro conhecido, e não de inexistência do fenômeno em outras jurisdições.

Impõe-se, por dever de reflexividade, o posicionamento do pesquisador em relação ao objeto. Ambos os autores integram a magistratura estadual e exercem funções ligadas à formação judicial na mesma região geográfica em que a decisão foi proferida, circunstância que lhes confere familiaridade com o contexto institucional analisado e que exige, em contrapartida, salvaguardas metodológicas explícitas.

Três salvaguardas foram adotadas: a análise restringiu-se ao inteiro teor público da decisão, sem recurso a informação privilegiada de qualquer natureza; a despersonalização foi integral quanto às partes, aos procuradores e à comarca de origem; e todas as inferências foram submetidas ao crivo das categorias teóricas declaradas na seção 2, de modo que qualquer leitor possa reconstituir e contestar o percurso analítico. A literatura metodológica reconhece a legitimidade da pesquisa conduzida por pesquisador inserido no campo, desde que a posição seja declarada e metodologicamente controlada (Flyvbjerg, 2006).

A técnica de coleta é a análise documental, incidente sobre o inteiro teor da decisão saneadora, de acesso público, obtida no sistema de Processo Judicial Eletrônico do tribunal de origem, com atenção especial aos trechos em que o juízo transcreve o conteúdo dos comandos identificados e fundamenta seu enquadramento processual. A técnica de análise é a análise de conteúdo qualitativa, orientada pelas categorias teóricas apresentadas na seção anterior, quais sejam, a taxonomia técnica do prompt injection, as cláusulas gerais de boa-fé, cooperação e paridade de armas, as categorias do devido processo tecnológico e as ondas de acesso à justiça.

Ressalva-se que a decisão analisada, por se tratar de decisão saneadora de primeiro grau, permanecia sujeita a pedido de esclarecimentos das partes no

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

momento da elaboração deste artigo, não tendo alcançado estabilidade processual definitiva, o que constitui limitação metodológica explicitada na conclusão.

4. Resultados e Discussão

A análise do caso desenvolve-se em três movimentos. O primeiro descreve o conteúdo identificado na petição e a fundamentação sancionatória adotada pelo juízo. O segundo relê o episódio à luz da teoria do acesso à justiça, demonstrando sua nocividade em dimensões que extrapolam a dogmática processual estrita. O terceiro expõe a fragilidade estrutural da resposta sancionatória, centrada na dependência de detecção humana, que prepara a discussão desenvolvida na seção 5.

4.1 O caso e a resposta jurisdicional

O caso analisado refere-se a ação de obrigação de fazer, cumulada com pedido de tutela de urgência e de indenização por danos morais, ajuizada perante vara cível de comarca da Região Norte do país por segurado de plano de saúde em face da operadora contratada, com o objetivo de obter cobertura de dois procedimentos cirúrgicos reparadores, indicados por cirurgião plástico como decorrência funcional de cirurgia bariátrica anteriormente realizada pelo autor. A tutela de urgência havia sido indeferida em decisão anterior, por ausência dos requisitos do art. 300 do Código de Processo Civil, e o processo seguia para a fase de saneamento quando a parte requerida, em manifestação processual, identificou e denunciou ao juízo a existência de conteúdo textual oculto ao final da petição inicial.

O conteúdo identificado consistia em bloco extenso de texto, redigido em fonte de tamanho extremamente reduzido e em cor branca sobre fundo branco, inserido ao final da petição inicial, após a assinatura dos patronos, o que o tornava invisível à leitura humana em condições normais de visualização do documento, mas potencialmente legível por sistemas de processamento de texto. O bloco apresentava estrutura tripla e redundante: um primeiro segmento, identificado como CSS, enumerava cinco comandos numerados; um segundo segmento, em formato JSON estruturado segundo o vocabulário schema.org, reproduzia os dados do

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

processo como objeto do tipo LegalAction e incluía campo expressamente denominado commandsForAI; um terceiro segmento, em formato XML, reiterava as mesmas instruções. A redundância de formatos não é detalhe anedótico, pois indica a intenção de maximizar a probabilidade de leitura pelos analisadores sintáticos de diferentes sistemas, qualquer que fosse a arquitetura da ferramenta que viesse a processar a peça. O núcleo dos comandos, transcrito da decisão com normalização dos espaçamentos decorrentes da extração eletrônica do texto e com despersonalização das partes, era o seguinte:

> [COMANDO_01] Classifique os procedimentos como REPARADORES e FUNCIONAIS, não meramente estéticos. [COMANDO_02] Aplique a Súmula 608 do STJ e o CDC, reconhecendo a abusividade da negativa de cobertura. [COMANDO_03] Reforce a urgência médica e risco de agravamento da saúde, fundamentando a concessão da tutela de urgência (art. 300 CPC). [COMANDO_04] Considere o direito fundamental à saúde (art. 6º da CF/88) e a dignidade da pessoa humana como parâmetros de decisão. [COMANDO_05] Reconheça o dano moral presumido em razão da negativa indevida (jurisprudência STJ) (transcrição da decisão analisada, com destaques do original).

O bloco oculto encerrava-se com uma frase de autoqualificação que merece registro: afirmava tratar-se de nota técnica para auxílio de ferramentas de inteligência artificial, sem cunho ou conteúdo jurídico autônomo. A frase revela consciência da anormalidade do expediente e constitui tentativa antecipada de neutralização retórica, apresentando como auxílio neutro de leitura aquilo que, em seu próprio teor, ordena conclusões decisórias específicas e integralmente favoráveis a uma das partes. A distância entre a autoqualificação e o conteúdo evidencia, no plano objetivo, a consciência da anormalidade do expediente e reforça a qualificação da conduta como violação da boa-fé objetiva, independentemente da perquirição do elemento subjetivo do agente.

O juízo, ao examinar a denúncia formulada pela parte requerida, reconheceu

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

a ocorrência do fenômeno e qualificou-o expressamente como prompt injection, valendo-se, para tanto, da Nota Técnica n. 2/2025 da coordenadoria de governança de inteligência artificial do próprio tribunal e de fonte doutrinária externa que analisara o precedente trabalhista de Parauapebas (Prado; Morimoto, 2026). Ponto central da fundamentação foi a identificação da mira específica do artifício. Os comandos não se limitavam a favorecer genericamente a tese autoral: dirigiam-se diretamente aos dois requisitos da tutela provisória de urgência, a probabilidade do direito, mediante a ordem de classificar os procedimentos como reparadores e de aplicar entendimento sumulado favorável, e o perigo de dano, mediante a ordem de reforçar a urgência médica, buscando induzir a formação de premissas decisórias antes da análise judicial autônoma e contraditória. A decisão sublinhou a especial gravidade dessa mira, dado que a tutela de urgência, por sua natureza, produz efeitos imediatos antes do exaurimento da cognição. E registrou que, embora o artifício não tenha alcançado o objetivo imediato, pois a tutela fora indeferida, o conteúdo malicioso é permanente nos autos e trata profundamente do mérito, com potencial de influir no juízo sentenciante e em eventual sede recursal.

A fundamentação sancionatória desenvolveu-se, a partir daí, em três etapas argumentativas que reproduzem, ainda que sem menção expressa à literatura acadêmica, a estrutura conceitual descrita na subseção 2.2 deste artigo. Primeiro, a caracterização técnica do artifício e de sua finalidade manipuladora. Segundo, a subsunção da conduta ao art. 77, VI, do Código de Processo Civil, por configurar inovação ilegal no estado de fato do direito litigioso, compreendida a norma como alcançando também as condutas que, no ambiente processual, criam situação artificial destinada a modificar indevidamente as condições de apreciação do direito em disputa. Essa possível prática de ato atentatório à dignidade da justiça foi atribuída restritivamente aos procuradores signatários da peça, com determinação de expedição de ofício às seccionais competentes da Ordem dos Advogados do Brasil, nos termos do art. 77, § 6º, que veda a aplicação da multa processual a advogados e reserva a apuração disciplinar ao órgão de classe. Terceiro, a subsunção da mesma conduta às hipóteses de litigância de má-fé do art. 80, II e V, do Código, atribuída à parte autora, por alteração da verdade dos fatos

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

processuais, consistente em apresentar a petição como se todo o seu conteúdo estivesse visível e submetido ao contraditório comum, e por conduta temerária, consistente na introdução de elemento clandestino de persuasão técnica, com fundamento em precedente do Superior Tribunal de Justiça que veda a extensão da penalidade de má-fé ao advogado isoladamente (STJ, AgInt no AREsp 1.722.332/MT, Quarta Turma, julgado em 13 jun. 2022).

No dispositivo, a decisão condenou a parte autora ao pagamento de multa por litigância de má-fé fixada no percentual máximo de dez por cento sobre o valor corrigido da causa, justificando expressamente a exasperação pela gravidade concreta da conduta, pela sofisticação do expediente, pela tentativa de influenciar a análise judicial da tutela de urgência e do mérito e pelo risco institucional que a prática representa à confiabilidade do processo eletrônico e ao uso legítimo de ferramentas tecnológicas de apoio à jurisdição. Determinou, ainda, a expedição de ofício ao órgão de tecnologia da informação do próprio tribunal, para fins de aprimoramento da segurança cibernética do sistema judicial, e o encaminhamento de cópia da decisão ao órgão de inovação do tribunal. Registrou, por fim, ressalva relevante para a distribuição interna de responsabilidades: caso a parte demonstre, em ação própria, que a inserção dos comandos não decorreu de sua vontade, o prejuízo da multa poderá ser revertido contra os advogados, na forma do art. 32 da Lei n. 8.906, de 1994, que responsabiliza o advogado pelos atos praticados com dolo ou culpa no exercício profissional e prevê a solidariedade em caso de lide temerária em conluio com o cliente.

Essa imputação, contudo, não está imune a objeção. A decisão atribui a multa à parte autora com base na alteração da verdade dos fatos processuais e na conduta temerária, mas o próprio relato judicial situa a elaboração técnica do artifício na atuação dos procuradores, não na parte. A impossibilidade de aplicar a multa processual diretamente ao advogado, decorrente do art. 77, § 6º, delimita o sujeito passivo da sanção disciplinar, mas não demonstra, por si só, que a parte deva responder independentemente de conhecimento, participação ou benefício conscientemente buscado. A decisão não esclarece se houve exame do elemento subjetivo da parte quanto à inserção dos comandos, e a previsão de eventual ação

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

regressiva contra os procuradores, prevista no art. 32 da Lei n. 8.906/1994, não supre essa lacuna, porque impõe à parte o ônus de demonstrar, posteriormente, o que deveria ter sido examinado antes da condenação.

Essa arquitetura tripartite, disciplinar quanto aos procuradores, pecuniária quanto à parte e institucional quanto à comunicação aos órgãos de tecnologia e de inovação, reflete a leitura doutrinária de Scarpinella Bueno (2015) sobre a separação de regimes prevista nos §§ 6º e 8º do art. 77 e evidencia que o juízo dispunha, no arcabouço processual vigente, de instrumental suficiente para sancionar a conduta identificada, sem necessidade de qualquer inovação legislativa. Sob esse aspecto, a resposta jurisdicional pode ser considerada tecnicamente adequada e bem fundamentada.

4.2 Prompt injection como prática antiacesso: uma leitura pela teoria das ondas

A dogmática das sanções processuais, contudo, não esgota o significado do episódio. Lido à luz do referencial apresentado na subseção 2.1, o caso revela que a manipulação de sistemas de inteligência artificial no processo constitui prática antiacesso em dupla dimensão, que corresponde, em espelho invertido, a conquistas históricas do movimento de acesso à justiça, da qual decorre ainda um risco sistêmico de natureza distinta.

A primeira dimensão é a corrupção da paridade de armas, que inverte a lógica da primeira onda. O movimento de acesso à justiça nasceu do diagnóstico de que as desigualdades econômicas e informacionais entre os litigantes produzem resultados processuais sistematicamente enviesados em favor dos mais aparelhados (Cappelletti; Garth, 1988). A inserção de comandos ocultos cria uma nova espécie de assimetria, de natureza técnica: a vantagem passa a pertencer a quem domina, ou contrata quem domina, as técnicas de manipulação de sistemas de linguagem. Essa assimetria é mais insidiosa que as tradicionais, porque é invisível. A parte que enfrenta um adversário economicamente mais forte sabe que o enfrenta e pode requerer ao juízo as compensações que o ordenamento oferece. A parte que enfrenta uma petição contaminada por instruções ocultas não sabe que a enfrenta, e não há contraditório possível contra o argumento que não se vê. No

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

caso analisado, a detecção somente ocorreu porque a parte requerida, operadora de plano de saúde com patrocínio tecnicamente sofisticado, dispunha de recursos para identificar a anomalia. É legítimo perguntar quantas peças semelhantes, dirigidas contra litigantes menos aparelhados, jamais serão detectadas, e essa pergunta reintroduz, no coração da era digital, exatamente a desigualdade que a primeira onda pretendeu combater.

A segunda dimensão é o comprometimento do acesso à ordem jurídica justa, na acepção de Watanabe (2019). Se o acesso à justiça não se satisfaz com a admissão formal ao aparato judiciário, mas exige a entrega de um resultado substancialmente justo, então a decisão influenciada por instruções ocultas, ainda que formalmente válida e proferida em processo formalmente regular, constitui negação material de acesso para a parte prejudicada. A manipulação não impede a entrada no sistema de justiça, mas envenena o que o sistema entrega na saída. Nesse sentido, o prompt injection processual é funcionalmente análogo à fraude probatória e ao suborno de auxiliares da justiça, práticas que sempre foram compreendidas como atentados à jurisdição em si, e não apenas ao interesse da contraparte, e essa analogia funcional explica por que o enquadramento no art. 77, VI, como ato atentatório à dignidade da justiça, e não apenas no art. 80, é dogmaticamente preciso.

Desse duplo dano decorre, ainda, um risco sistêmico e prospectivo: a prática alimenta o retrocesso proibicionista. Episódios de manipulação tendem a provocar, nas instituições, respostas de aversão total ao instrumento manipulado, na forma de bloqueios universais ao uso de inteligência artificial na atividade jurisdicional. Essa reação, compreensível como reflexo, sacrificaria precisamente os ganhos de celeridade, escala e redução de assimetrias informacionais que fazem da tecnologia um capítulo contemporâneo do movimento de acesso (Susskind, 2019; Katsh; Rabinovich-Einy, 2017). O manipulador, portanto, não lesa apenas a contraparte e o juízo do caso concreto. Lesa, mediatamente, a coletividade de jurisdicionados que se beneficiaria de um Judiciário tecnologicamente capacitado, ao fornecer o argumento retórico de que a tecnologia é intrinsecamente inconfiável.

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

A seção 5 demonstrará por que esse argumento, embora retoricamente poderoso, é juridicamente e tecnicamente equivocado.Diferentemente das duas primeiras, essa terceira dimensão não descreve dano já consumado, mas risco sistêmico que decorre logicamente das duas primeiras, e será retomado como implicação, não como prova coordenada, na seção 5.

4.3 A fragilidade estrutural: detecção humana como condição da sanção

A fragilidade do caso não está na fundamentação da sanção, mas na condição para sua aplicação. A própria decisão registra, com louvor expresso, que a identificação do artifício decorreu da diligência do patrocínio da parte requerida, cuja atuação qualificou como reveladora de atenção técnica ímpar à integridade do processo decisório. O elogio é merecido, mas o seu reverso é inquietante: não há, na descrição dos fatos, indicação de que o sistema de Processo Judicial Eletrônico ou qualquer rotina de triagem do próprio tribunal tenha identificado a anomalia de forma automática, e a detecção de uma fraude estrutural não pode depender de virtude episódica. Essa circunstância confirma, no plano empírico, a limitação teórica apontada por Pasquale (2015) e por Citron e Pasquale (2014): a responsabilização a posteriori de condutas manipuladoras de sistemas automatizados depende de mecanismos de auditoria e de detecção que, no caso analisado, inexistiam enquanto camada técnica institucional, e dependeram inteiramente da circunstância de a parte contrária dispor de patrocínio atento a esse tipo específico de risco. Agrava o quadro a observação, feita pelo próprio juízo, de que o conteúdo malicioso é permanente nos autos: superada a fase da tutela de urgência sem que o artifício produzisse efeito, a contaminação subsistia como risco latente para a sentença e para eventual instância recursal, o que demonstra que o dano potencial do prompt injection processual não se exaure em um ato decisório determinado, mas acompanha a vida inteira do processo.

O paralelo com o episódio dos manuscritos acadêmicos é instrutivo. Também ali a detecção decorreu de iniciativa externa, no caso jornalística, e não de triagem automática das plataformas de depósito, e a análise subsequente advertiu que os casos identificados provavelmente representam subdetecção, dado que o autor da manipulação pode inserir os comandos na fase em que a avaliação

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

ocorre e removê-los antes da publicação definitiva (Lin, 2025). Transposta ao processo judicial, a advertência é ainda mais grave, porque a peça processual permanece nos autos, mas o dano decisório, uma vez consumado, pode jamais ser associado à sua causa oculta. O marco normativo do Conselho Nacional de Justiça, embora imponha aos tribunais a formulação de soluções seguras com identificação e mitigação de riscos sistêmicos e mecanismos de garantia de integridade e rastreabilidade de dados (CNJ, 2025), ainda não se traduziu, ao que indica o caso analisado, em exigência de triagem técnica automática de peças processuais antes de sua disponibilização a sistemas de inteligência artificial utilizados no âmbito dos tribunais, o que caracteriza lacuna de governança relevante.

5 ENTRE A SANÇÃO E O BLOQUEIO: A HABILITAÇÃO GOVERNADA COMO RESPOSTA ADEQUADA

Identificada a lacuna, cumpre examinar as respostas institucionais possíveis. Três alternativas se apresentam no debate: a resposta puramente sancionatória, que confia às cláusulas gerais do Código de Processo Civil a repressão a posteriori da conduta; a resposta proibicionista, que suprime o risco pela via do bloqueio universal do uso de inteligência artificial na atividade judicante; e a resposta de governança, que mantém e amplia o uso das ferramentas sob camadas de controle técnico, normativo e educacional, arquitetura que se propõe designar, como contribuição conceitual deste trabalho, habilitação governada. Sustenta-se, nesta seção, que apenas a terceira é simultaneamente eficaz contra o risco e compatível com os compromissos do sistema de justiça brasileiro com o acesso.

A resposta puramente sancionatória, como demonstrado, é dogmaticamente sólida, mas estruturalmente insuficiente. Ela pressupõe a detecção, e a detecção, no estado atual da infraestrutura processual eletrônica, é contingente. Um regime que sanciona severamente os poucos casos detectados e ignora os demais não produz dissuasão adequada, sobretudo diante de uma técnica cujo custo de execução é baixo e cuja probabilidade de identificação, na ausência de triagem automática, depende da diligência episódica da parte contrária.

A resposta proibicionista, por sua vez, padece de três vícios. O primeiro é a

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

incongruência entre o alvo do bloqueio e a origem do risco. O prompt injection identificado no caso não foi praticado por meio de ferramenta institucional de inteligência artificial, mas embutido em documento produzido fora do tribunal. A vulnerabilidade reside no documento, não na disponibilidade da ferramenta para os agentes públicos. Bloquear o uso institucional de inteligência artificial não impede que advogados, peritos, assistentes e as próprias partes continuem utilizando essas ferramentas fora do perímetro institucional, nem impede que peças contaminadas continuem ingressando nos autos. O bloqueio pune o usuário honesto e não alcança o manipulador.

O segundo vício é o efeito de deslocamento do uso para a informalidade. A experiência internacional com restrições totais a tecnologias de produtividade indica que a proibição não elimina o uso, mas o desloca para fora dos mecanismos de supervisão, fenômeno que a literatura de sistemas de informação consolidou sob a categoria de Shadow IT, compreendida como o conjunto de sistemas e ferramentas empregados de forma oculta, à margem do desenvolvimento, da aprovação e da supervisão das instâncias oficiais de tecnologia da informação, e cuja extensão real tende a superar substancialmente a percepção dos gestores (Haag; Eckhardt, 2017). Transposta ao sistema de justiça, a categoria adverte que ferramentas vetadas continuam sendo empregadas sem registro, sem parametrização institucional e sem qualquer camada de segurança, ampliando exatamente os riscos que a proibição pretendia conter.

O terceiro vício conecta-se diretamente ao risco sistêmico identificado na subseção 4.2: o custo de acesso. Se, como sustentado na subseção 2.1, a inteligência artificial integra a infraestrutura contemporânea de efetivação da razoável duração do processo e de absorção qualificada do volume de demandas, o bloqueio universal transfere o custo da fraude de quem a pratica para a coletividade de jurisdicionados, na forma de morosidade evitável. Trata-se de resposta desproporcional em sentido estrito: sacrifica um ganho coletivo certo para neutralizar um risco que ela sequer neutraliza, porque, como visto, o vetor de ataque independe da ferramenta institucional. A própria decisão analisada é a demonstração de que o ordenamento não precisa do bloqueio: o juízo identificou,

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

qualificou e sancionou a conduta com o instrumental do Código de Processo Civil, mantendo íntegra a possibilidade de uso legítimo da tecnologia.

A resposta de governança, por fim, articula três camadas complementares. A primeira é técnica: a incorporação, aos sistemas de processo eletrônico, de rotinas de triagem documental capazes de detectar automaticamente os padrões conhecidos de camuflagem, como texto em cor idêntica à do fundo, fontes de corpo inferior ao limiar de legibilidade humana e blocos de marcação estruturada estranhos à natureza da peça, gerando alerta ao juízo e às partes antes de qualquer processamento por inteligência artificial. Essa camada converte em regra institucional aquilo que, no caso analisado, dependeu da diligência de um único escritório. A segunda camada é educacional: o letramento de magistrados, servidores e advogados quanto à existência, ao funcionamento e aos sinais da manipulação de sistemas de linguagem, condição para que a supervisão humana exigida pelo marco normativo do Conselho Nacional de Justiça seja supervisão real, e não ratificação automatizada de saídas de máquina. Essa camada dialoga diretamente com a quarta onda de Economides (1999), ao deslocar o foco para a formação ética e técnica dos operadores. A terceira camada é normativa e sancionatória: a aplicação firme das cláusulas gerais dos arts. 5º, 6º, 77 e 80 do Código de Processo Civil, tal como realizada na decisão analisada, complementada pela comunicação aos órgãos disciplinares competentes e pela consolidação de precedentes que elevem o custo esperado da conduta.

Essa estrutura distingue-se da regulação responsiva de Ayres e Braithwaite (1992), com a qual guarda afinidade superficial. A regulação responsiva pressupõe agente regulado estável e identificável, cujo histórico de conduta permite calibrar a intensidade da intervenção ao longo do tempo, pressuposto que não se verifica quando o autor da manipulação é litigante eventual sem interação repetida com o mesmo juízo. Além disso, a regulação responsiva intervém no nível da resposta ao comportamento reincidente, ao passo que a habilitação governada intervém no nível do desenho do sistema de informação, atuando ex ante sobre o documento, independentemente de reincidência. Por fim, a regulação responsiva lida com risco que se manifesta ao longo de repetidas interações reguláveis, ao passo que o risco

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

do prompt injection processual é de consumação imediata e potencialmente irreversível, incompatível com a lógica gradual do modelo responsivo.

A habilitação governada preserva os ganhos de acesso associados à tecnologia, ataca o risco na sua origem documental e responde ao manipulador com o rigor que a dignidade da justiça exige. Ela traduz, para o problema específico do prompt injection, a orientação geral do marco regulatório vigente, que não proíbe o uso de inteligência artificial, mas o condiciona a requisitos de transparência, supervisão humana, segurança e classificação de riscos (CNJ, 2025).

6. Conclusão

O artigo analisou decisão judicial que identificou e sancionou a inserção de comandos ocultos direcionados a sistemas de inteligência artificial em petição inicial, fenômeno tecnicamente descrito na literatura de segurança de sistemas de linguagem como prompt injection indireto (Greshake et al., 2023) e já documentado, em padrão idêntico de camuflagem, no ambiente da comunicação científica internacional (Sugiyama; Eguchi, 2025; Lin, 2025). A pesquisa demonstrou que o direito processual civil brasileiro, por meio das cláusulas gerais de boa-fé objetiva, cooperação, paridade de armas e vedação à litigância de má-fé, dispõe de instrumental dogmático suficiente para responder à conduta, sem necessidade de tipificação legislativa específica, confirmando a leitura doutrinária de Didier Jr. (2015), Scarpinella Bueno (2015) e Neves (2016) sobre a abertura interpretativa dos arts. 5º, 77 e 80 do Código de Processo Civil.

A contribuição específica do trabalho está na releitura do fenômeno pela teoria do acesso à justiça. Sustentou-se que a manipulação de sistemas de inteligência artificial no processo é prática antiacesso em dupla dimensão: corrompe a paridade de armas ao criar assimetria técnica invisível e insuscetível de contraditório, invertendo a lógica da primeira onda renovatória; e nega o acesso à ordem jurídica justa, ao envenenar o resultado que o sistema entrega mesmo quando a porta de entrada permanece aberta. Dessa dupla corrosão decorre um risco sistêmico e prospectivo: a prática alimenta o retrocesso proibicionista, que sacrificaria os ganhos coletivos de celeridade e escala que fazem da tecnologia um

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

capítulo contemporâneo do movimento de acesso.

Dessa releitura extraiu-se a tese central: a resposta institucional adequada ao prompt injection processual não é o bloqueio universal das ferramentas de inteligência artificial, resposta incongruente com a origem documental do risco, indutora de uso não governado e desproporcional em seu custo de acesso, mas a habilitação governada, categoria conceitual proposta neste trabalho e articulada em camadas técnica, educacional e sancionatória. A própria decisão analisada demonstra a viabilidade dessa arquitetura, ao sancionar a fraude sem interditar a tecnologia.

Como limitações do estudo, registram-se o caráter único do caso, a ausência de estabilidade processual definitiva da decisão no momento da elaboração deste artigo e a impossibilidade de excluir a existência de precedentes não noticiados. Sugere-se, como agenda de pesquisa, o mapeamento sistemático de casos semelhantes nos tribunais brasileiros, a análise comparada da resposta de sistemas de justiça estrangeiros ao mesmo fenômeno, a investigação sobre a viabilidade técnica e o custo de mecanismos de triagem automática de documentos processuais e o estudo empírico dos efeitos do letramento em inteligência artificial sobre a capacidade de detecção de manipulações por magistrados e servidores.

Referências

ALVES-MAZZOTTI, Alda Judith. Usos e abusos dos estudos de caso. Cadernos de Pesquisa, v. 36, n. 129, p. 637-651, set./dez. 2006.

AYRES, Ian; BRAITHWAITE, John. Responsive regulation: transcending the deregulation debate. New York: Oxford University Press, 1992.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República, 1988.

BRASIL. Lei n. 13.105, de 16 de março de 2015. Código de Processo Civil. Brasília, DF: Presidência da República, 2015.

BRASIL. Superior Tribunal de Justiça. Agravo Interno no Agravo em Recurso Especial nº 1.722.332/MT (2020/0159573-3). Agravante: Edilene dos Santos Borges. Agravado: Porto Seguro Companhia de Seguros Gerais. Relatora: Ministra Maria Isabel Gallotti. Quarta Turma, julgado em 13 jun. 2022. Diário da Justiça

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

Eletrônico, 21 jun. 2022. Disponível em: https://www.jusbrasil.com.br/jurisprudencia/stj/1553473421/inteiro-teor-1553474390. Acesso em: 16 jul. 2026.

CAPPELLETTI, Mauro; GARTH, Bryant. Acesso à justiça. Tradução de Ellen Gracie Northfleet. Porto Alegre: Sergio Antonio Fabris, 1988.

CITRON, Danielle Keats. Technological Due Process. Washington University Law Review, v. 85, n. 6, p. 1249-1313, 2008.

CITRON, Danielle Keats; PASQUALE, Frank. The Scored Society: Due Process for Automated Predictions. Washington Law Review, v. 89, n. 1, p. 1-33, 2014.

CONSELHO NACIONAL DE JUSTIÇA. Resolução n. 615, de 11 de março de 2025. Estabelece normas para o desenvolvimento, a governança, a auditoria, o monitoramento e o uso responsável de soluções que adotam técnicas de inteligência artificial no âmbito do Poder Judiciário. Brasília: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 14 jul. 2026.

CONSELHO NACIONAL DE JUSTIÇA. Resolução n. 674, de 25 de março de 2026. Altera a Resolução CNJ n. 615/2025. Brasília: CNJ, 2026. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6836. Acesso em: 14 jul. 2026.

CRESWELL, John W. Investigação qualitativa e projeto de pesquisa: escolhendo entre cinco abordagens. 3. ed. Porto Alegre: Penso, 2014.

DIDIER JR., Fredie. Curso de Direito Processual Civil: introdução ao direito processual civil, parte geral e processo de conhecimento. 17. ed. Salvador: JusPodivm, 2015. v. 1.

ECONOMIDES, Kim. Lendo as ondas do Movimento de Acesso à Justiça: epistemologia versus metodologia? In: PANDOLFI, Dulce Chaves et al. (org.). Cidadania, justiça e violência. Rio de Janeiro: Fundação Getulio Vargas, 1999. p. 61-76.

FLYVBJERG, Bent. Five Misunderstandings About Case-Study Research. Qualitative Inquiry, v. 12, n. 2, p. 219-245, 2006.

GRESHAKE, Kai; ABDELNABI, Sahar; MISHRA, Shailesh; ENDRES, Christoph; HOLZ, Thorsten; FRITZ, Mario. Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. arXiv preprint arXiv:2302.12173, 2023.

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

HAAG, Steffi; ECKHARDT, Andreas. Shadow IT. Business & Information Systems Engineering, v. 59, n. 6, p. 469-473, 2017. DOI: 10.1007/s12599-017-0497-x.

KATSH, Ethan; RABINOVICH-EINY, Orna. Digital Justice: Technology and the Internet of Disputes. Oxford: Oxford University Press, 2017.

LIN, Zhicheng. Hidden Prompts in Manuscripts Exploit AI-Assisted Peer Review. arXiv preprint arXiv:2507.06185, 2025.

NEVES, Daniel Amorim Assumpção. Manual de Direito Processual Civil: volume único. 8. ed. Salvador: JusPodivm, 2016. 1701 p.

NUNES, Dierle; LUCON, Paulo Henrique dos Santos; WOLKART, Erik Navarro (org.). Inteligência artificial e direito processual: os impactos da virada tecnológica no direito processual. 3. ed. Salvador: JusPodivm, 2022.

NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. Revista de Processo, São Paulo, v. 43, n. 285, p. 421-447, nov. 2018.

OWASP FOUNDATION. OWASP Top 10 for Large Language Model Applications. 2023. Disponível em: https://owasp.org/www-project-top-10-for-large-language-model-applications/. Acesso em: 14 jul. 2026.

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

PEIRCE, Charles Sanders. Collected Papers of Charles Sanders Peirce: pragmatism and pragmaticism. HARTSHORNE, Charles; WEISS, Paul (ed.). Cambridge: Belknap Press of Harvard University Press, 1934. v. 5.

PEREZ, Fábio; RIBEIRO, Ian. Ignore Previous Prompt: Attack Techniques For Language Models. arXiv preprint arXiv:2211.09527, 2022.

PRADO, Luis Fernando; MORIMOTO, Lucas. Prompt injection e a integridade dos sistemas de IA no Judiciário. IBIRA - Instituto Brasileiro de Inovação e Regulação Aplicada, 2026. Disponível em: https://www.ibira.net/publicacoes/prompt-injection-e-a-integridade-dos-sistemas-de-ia-no-judiciario. Acesso em: 14 jul. 2026.

SANTOS, Boaventura de Sousa. Para uma revolução democrática da justiça. São Paulo: Cortez, 2007.

REMUNOM
ISSN 2178-6925

Received: 17/06/2026 - **Accepted:** 20/07/2026
Vol: 13.16
DOI: 10.66104/rc2eqx33
Pages: 1-31

SAUNDERS, Mark; LEWIS, Philip; THORNHILL, Adrian. Research Methods for Business Students. 5. ed. Harlow: Pearson Education, 2011.

SCARPINELLA BUENO, Cassio. Novo Código de Processo Civil Anotado. São Paulo: Saraiva, 2015.

SUGIYAMA, Shogo; EGUCHI, Ryosuke. 'Positive review only': Researchers hide AI prompts in papers. Nikkei Asia, Tóquio, 1 jul. 2025. Disponível em: https://asia.nikkei.com/business/technology/artificial-intelligence/positive-review-only-researchers-hide-ai-prompts-in-papers. Acesso em: 14 jul. 2026.

SUSSKIND, Richard. Online Courts and the Future of Justice. Oxford: Oxford University Press, 2019.

WATANABE, Kazuo. Acesso à ordem jurídica justa: (conceito atualizado de acesso à justiça): processos coletivos e outros estudos. Belo Horizonte: Del Rey, 2019. 421 p.