

IX ENCONTRO VIRTUAL DO CONPEDI

**DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS
IV**

Todos os direitos reservados e protegidos. Nenhuma parte destes anais poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

Diretoria - CONPEDI

Presidente - Profa. Dra. Samyra Haydêe Dal Farra Naspolini - FMU - São Paulo

Diretor Executivo - Prof. Dr. Orides Mezzaroba - UFSC - Santa Catarina

Vice-presidente Norte - Prof. Dr. Jean Carlos Dias - Cesupa - Pará

Vice-presidente Centro-Oeste - Prof. Dr. José Querino Tavares Neto - UFG - Goiás

Vice-presidente Sul - Prof. Dr. Leonel Severo Rocha - Unisinos - Rio Grande do Sul

Vice-presidente Sudeste - Profa. Dra. Rosângela Lunardelli Cavallazzi - UFRJ/PUCRio - Rio de Janeiro

Vice-presidente Nordeste - Prof. Dr. Raymundo Juliano Feitosa - UNICAP - Pernambuco

Representante Discente: Prof. Dr. Abner da Silva Jaques - UPM/UNIGRAN - Mato Grosso do Sul

Conselho Fiscal:

Prof. Dr. José Filomeno de Moraes Filho - UFMA - Maranhão

Prof. Dr. Caio Augusto Souza Lara - SKEMA/ESDHC/UFMG - Minas Gerais

Prof. Dr. Valter Moura do Carmo - UFRSA - Rio Grande do Norte

Prof. Dr. Fernando Passos - UNIARA - São Paulo

Prof. Dr. Ednilson Donisete Machado - UNIVEM/UENP - São Paulo

Secretarias

Relações Institucionais:

Prof. Dra. Claudia Maria Barbosa - PUCPR - Paraná

Prof. Dr. Heron José de Santana Gordilho - UFBA - Bahia

Profa. Dra. Daniela Marques de Moraes - UNB - Distrito Federal

Comunicação:

Prof. Dr. Robison Tramontina - UNOESC - Santa Catarina

Prof. Dr. Liton Lanes Pilau Sobrinho - UPF/Univali - Rio Grande do Sul

Prof. Dr. Lucas Gonçalves da Silva - UFS - Sergipe

Relações Internacionais para o Continente Americano:

Prof. Dr. Jerônimo Siqueira Tybusch - UFSM - Rio Grande do Sul

Prof. Dr. Paulo Roberto Barbosa Ramos - UFMA - Maranhão

Prof. Dr. Felipe Chiarello de Souza Pinto - UPM - São Paulo

Relações Internacionais para os demais Continentes:

Profa. Dra. Gina Vidal Marcilio Pompeu - UNIFOR - Ceará

Profa. Dra. Sandra Regina Martini - UNIRITTER / UFRGS - Rio Grande do Sul

Profa. Dra. Maria Claudia da Silva Antunes de Souza - UNIVALI - Santa Catarina

Educação Jurídica

Profa. Dra. Viviane Coêlho de Séllos Knoerr - Unicuritiba - PR

Prof. Dr. Rubens Beçak - USP - SP

Profa. Dra. Livia Gaigher Bosio Campello - UFMS - MS

Eventos:

Prof. Dr. Yuri Nathan da Costa Lannes - FDF - São Paulo

Profa. Dra. Norma Sueli Padilha - UFSC - Santa Catarina

Prof. Dr. Juraci Mourão Lopes Filho - UNICHRISTUS - Ceará

Comissão Especial

Prof. Dr. João Marcelo de Lima Assafim - UFRJ - RJ

Profa. Dra. Maria Creusa De Araújo Borges - UFPB - PB

Prof. Dr. Antônio Carlos Diniz Murta - Fumec - MG

Prof. Dr. Rogério Borba - UNIFACVEST - SC

D597

Direito, governança e novas tecnologias IV [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Irineu Francisco Barreto Junior; Diogo de Almeida Viana dos Santos – Florianópolis: CONPEDI, 2026.

Inclui bibliografia

ISBN: 978-65-5274-588-0

Modo de acesso: www.conpedi.org.br em publicações

Tema: Constituição, Processo e Justiça Social: o papel da jurisdição constitucional na efetivação de direitos

2. Direito. 3. Governança e novas tecnologias. IX Encontro Virtual do CONPEDI (2: 2026: Florianópolis, Brasil).

CDU: 34



IX ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS IV

Apresentação

Os trabalhos que integram esta publicação foram apresentados no Grupo de Trabalho “Direito, Governança e Novas Tecnologias IV”, realizado virtualmente no dia 25 de junho de 2026, como parte da programação do IX Encontro Virtual do Conselho Nacional de Pesquisa e Pós-graduação em Direito (CONPEDI). Esta edição do evento reafirma o compromisso histórico desta sociedade científica, que há décadas promove a excelência na pesquisa jurídica, adaptando-se às novas realidades de interação digital para garantir a continuidade do diálogo acadêmico em nível nacional e internacional.

O presente Grupo de Trabalho consolidou-se como um espaço privilegiado de interlocução, destacando-se não apenas pela densidade teórica dos trabalhos expostos, mas também pelo perfil de seus participantes, composto por professores pesquisadores, doutores, mestres e alunos de pós-graduação, que trouxeram o frescor de suas investigações institucionais para o debate coletivo. A dinâmica das apresentações, conduzida em ambiente virtual, propiciou uma interação instigante entre os autores e o público, permitindo que a pesquisa jurídica fosse discutida sob o prisma da transdisciplinaridade e do enfrentamento dos desafios contemporâneos.

A temática central deste GT, focada na governança e nas novas tecnologias, reflete a inquietude necessária diante das constantes transformações da sociedade da informação. Cientes da complexidade desses fenômenos, os pesquisadores buscaram articular respostas jurídicas para uma realidade que se apresenta em fluxo permanente, onde a técnica e o Direito se entrelaçam. Nesse sentido, as discussões perpassaram por eixos fundamentais que estruturam o debate acadêmico atual, tais como:

Tecnologias Emergentes e Sociedade: A investigação sobre os impactos da cibercultura na esfera pública, o uso da tecnologia como ferramenta de persuasão e os desafios éticos impostos pela biotecnologia e pela segurança pública.

Os trabalhos selecionados para esta coletânea são o resultado de pesquisas rigorosas desenvolvidas em diversos Programas de Pós-graduação em Direito do país. Ressalte-se que todos os artigos foram submetidos a um criterioso processo de seleção, por meio do sistema de dupla avaliação cega por pares (double blind peer review), garantindo o rigor científico e a adequação temática aos propósitos deste Grupo de Trabalho.

Ao apresentar estes anais ao público, os coordenadores manifestam a expectativa de que estas reflexões contribuam efetivamente para o fortalecimento do diálogo interdisciplinar e para o avanço da ciência jurídica brasileira. O objetivo é oferecer aos leitores um panorama crítico e atualizado sobre como o Direito pode e deve mediar as relações em um mundo cada vez mais pautado pela inovação tecnológica, assegurando sempre a prevalência da dignidade humana e dos valores democráticos.

Desejamos a todos uma leitura proveitosa e inspiradora.

Os Coordenadores

Prof. Dr. Irineu Francisco Barreto Junior (Mestrado em Direito das Faculdades Metropolitanas Unidas - FMU-SP)

Prof. Dr. Diogo de Almeida Viana dos Santos (Universidade Estadual do Maranhão; Universidade CEUMA-PPG em Direito e Afirmação de Vulneráveis)

ENTRE O VISÍVEL E O INVISÍVEL: PROMPT INJECTION E A MANIPULAÇÃO ALGORÍTMICA NO PROCESSO JUDICIAL ELETRÔNICO

BETWEEN THE VISIBLE AND THE INVISIBLE: PROMPT INJECTION AND ALGORITHMIC MANIPULATION IN ELECTRONIC JUDICIAL PROCEEDINGS

Gilberto Batista Santos ¹

Resumo

O presente artigo tem por objetivo analisar os riscos produzidos pela utilização de técnicas de prompt injection invisível no contexto do processo judicial eletrônico, especialmente diante da crescente incorporação de sistemas de inteligência artificial generativa ao funcionamento do Poder Judiciário brasileiro. O trabalho parte da análise do processo de digitalização da atividade jurisdicional, da evolução da inteligência artificial generativa e do conceito de prompt, até alcançar o fenômeno do prompt injection invisível e suas implicações para a segurança informacional do processo judicial eletrônico. Há de se considerar que a crescente utilização institucional de sistemas generativos exige não apenas regulamentação tecnológica, mas também mecanismos capazes de preservar a estabilidade dos ambientes processuais digitais contemporâneos. Para tanto, aplica-se à pesquisa abordagem qualitativa, exploratória e analítica, utilizando como principal procedimento metodológico a análise documental e empírica da sentença proferida nos autos da Ação Trabalhista nº 0001062-55.2025.5.08.0130. A pesquisa encontra-se vinculada ao projeto de doutorado “Decisões Jurídicas Produzidas por Inteligência Artificial Generativa: Modelagem Estrutural e de Rastreabilidade no Processo Decisório”, desenvolvido no Programa de Pós-Graduação em Difusão do Conhecimento (PPGDC), bem como às atividades do Grupo de Estudo em Direito Digital e Inteligência Artificial (GEDDAI) e do projeto “404 – Núcleo de Observação Tecnopolítica sobre Sistemas Digitais”.

Palavras-chave: Inteligência artificial generativa, Prompt injection invisível, Processo judicial eletrônico, Ecossistemas processuais digitais, Governança algorítmica

but also mechanisms capable of preserving the stability of contemporary digital procedural environments. To this end, the research adopts a qualitative, exploratory, and analytical approach, using as its main methodological procedure the documentary and empirical analysis of the judgment rendered in Labor Lawsuit No. 0001062-55.2025.5.08.0130. The research is linked to the doctoral project entitled “Legal Decisions Produced by Generative Artificial Intelligence: Structural and Traceability Modeling in the Decision-Making Process”, developed within the Graduate Program in Knowledge Diffusion (PPGDC), as well as to the activities of the Study Group on Digital Law and Artificial Intelligence (GEDDAI) and the project “404 – Technopolitical Observatory on Digital Systems”.

Keywords/Palabras-claves/Mots-clés: Generative artificial intelligence, Invisible prompt injection, Electronic judicial proceedings, Digital procedural ecosystems, Algorithmic governance

INTRODUÇÃO

No século XXI, a sociedade presencia transformações capazes de alterar profundamente a forma como os indivíduos produzem conhecimento, estabelecem relações sociais e interagem com as instituições estatais. O avanço tecnológico observado nas últimas décadas demonstra que a humanidade passou a operar em um ambiente marcado pela intensa circulação de informações e pela crescente incorporação de sistemas computacionais às atividades cotidianas. Nesse sentido pode-se dizer que a tecnologia passou a integrar diretamente a dinâmica da vida social. Hoje trabalhamos, consumimos, nos comunicamos e até construímos nossas relações pessoais em ambientes digitais.

Diante disso o próprio acesso à justiça começa a sofrer impactos diretos dessas transformações tecnológicas. O processo judicial, historicamente estruturado em autos físicos e práticas presenciais passa gradativamente a desenvolver-se em sistemas processuais digitais compostos por sujeitos processuais, documentos eletrônicos e sistemas automatizados de processamento informacional.

Vale dizer que o documento processual eletrônico deixa de existir apenas como objeto de leitura humana para funcionar também como entrada informacional destinada à interpretação algorítmica. O que antes era pensado apenas como ambiente digital destinado à tramitação processual hoje passa a funcionar também como espaço de interação entre linguagem humana e sistemas computacionais capazes de interpretar textos, organizar documentos e estruturar informações jurídicas.

Em outras palavras, o processo eletrônico deixa de ser apenas uma ferramenta tecnológica do Judiciário e passa a integrar uma lógica muito mais complexa marcada pela mediação entre operadores do Direito e modelos algorítmicos de linguagem. Essa alteração da dinâmica jurisdicional intensificou-se significativamente com a expansão do Processo Judicial Eletrônico (PJe), da virtualização dos atos processuais e da crescente utilização de sistemas de inteligência artificial no âmbito do Poder Judiciário brasileiro.

Atualmente existem ferramentas capazes de interpretar linguagem natural, estruturar argumentos, resumir documentos e até auxiliar atividades relacionadas à produção de textos jurídicos. Não estamos afirmando que a utilização de inteligência artificial seja incompatível com o devido processo legal entretanto, é evidente que essas tecnologias passaram a produzir novos desafios relacionados à segurança jurídica, integridade documental e confiabilidade das estruturas decisórias contemporâneas.

Nesse diapasão, a crescente utilização institucional de inteligência artificial faz emergir discussões relacionadas à rastreabilidade decisória, preservação das garantias fundamentais e

proteção da própria integridade informacional dos ambientes processuais digitais. Diferentemente das vulnerabilidades tradicionais do processo eletrônico algumas dessas novas formas de interferência não atuam diretamente sobre a infraestrutura técnica dos sistemas, mas sobre a própria linguagem utilizada nesses ambientes. O problema deixa de estar apenas no código e passa a ocupar um espaço mais complexo, o da comunicação invisível estabelecida entre documento e máquina.

No âmbito do Poder Judiciário brasileiro essa preocupação ganhou maior relevância com a edição da Resolução nº 615/2025 do Conselho Nacional de Justiça (CNJ), que estabelece diretrizes para utilização de inteligência artificial generativa na atividade jurisdicional. Na mesma linha, o Plano Brasileiro de Inteligência Artificial (2024–2028), denominado “IA para o Bem de Todos”, reconhece a necessidade de criação de mecanismos voltados à governança, monitoramento e utilização responsável da inteligência artificial em estruturas públicas e ambientes decisórios.

Entretanto o avanço desses sistemas também faz surgir novas formas de manipulação relacionadas ao funcionamento dos modelos de linguagem de grande escala (*Large Language Models* – LLMs). Esses sistemas podem ser influenciados mediante comandos linguísticos destinados a alterar respostas, modificar interpretações ou contornar limitações internas do modelo.

É justamente nesse ambiente de crescente dependência linguística dos sistemas generativos que surge o denominado *prompt injection* invisível, técnica caracterizada pela inserção de comandos ocultos em documentos digitais mediante utilização de elementos não perceptíveis ao leitor humano, mas interpretáveis por sistemas algorítmicos de processamento textual.

A relevância jurídica do tema surge após uma decisão proferida pela Justiça do Trabalho da 8ª Região, em que o magistrado identificou a inserção de comandos ocultos em peça processual eletrônica destinados à manipulação do sistema de inteligência artificial utilizado na atividade jurisdicional. O caso revela uma alteração significativa na própria lógica contemporânea do documento processual eletrônico.

Em um processo físico tradicional um texto oculto em fonte branca dificilmente produziria qualquer consequência prática relevante. Entretanto, em ecossistemas processuais mediados por inteligência artificial o mesmo comando invisível pode funcionar como mecanismo de direcionamento algorítmico, estabelecendo uma espécie de comunicação oculta entre documento e máquina.

A discussão proposta pelo presente artigo vincula-se às atividades desenvolvidas pelo Grupo de Estudo em Direito Digital e Inteligência Artificial (GEDDAI). A pesquisa também dialoga diretamente com o projeto de doutorado intitulado “Decisões Jurídicas Produzidas por Inteligência Artificial Generativa: Modelagem Estrutural e de Rastreabilidade no Processo Decisório”, desenvolvido no âmbito do Programa de Pós-Graduação em Difusão do Conhecimento (PPGDC). A escolha do tema decorre ainda da necessidade de compreender como sistemas generativos podem impactar estruturas decisórias historicamente construídas para interação exclusivamente humana.

Metodologicamente a pesquisa adota abordagem qualitativa, exploratória e analítica utilizando como procedimento central a análise documental da sentença proferida nos autos da Ação Trabalhista nº 0001062-55.2025.5.08.0130, em trâmite perante a 3ª Vara do Trabalho de Parauapebas, no Estado do Pará. Na referida decisão o magistrado reconhece expressamente a utilização de técnica de *prompt injection* invisível mediante inserção de comandos ocultos em uma petição.

Partindo dessas premissas o presente artigo busca analisar os riscos decorrentes da utilização de técnicas de *prompt injection* invisível no contexto do processo judicial eletrônico, investigando de que forma essas estratégias de manipulação desafiam pressupostos tradicionais de boa-fé processual, integridade documental e confiabilidade da utilização institucional de sistemas de inteligência artificial generativa.

A DIGITALIZAÇÃO DO PROCESSO JUDICIAL E A FORMAÇÃO DE ECOSISTEMAS PROCESSUAIS DIGITAIS

Durante grande parte da história do processo brasileiro a atividade jurisdicional esteve vinculada aos autos físicos. A própria tramitação processual dependia da movimentação material de documentos, da atuação presencial dos sujeitos processuais e da capacidade operacional das estruturas cartorárias responsáveis pela gestão física dos autos, o tempo do processo confundia-se com o deslocamento entre cartórios, gabinetes, escritórios e tribunais, o processo era lento, burocrático e dependente da própria limitação física. Entretanto, com o crescimento exponencial da litigiosidade tornou impossível a manutenção do modelo jurisdicional de autos físico.

Nesse sentido, destaca Nathaly Campitelli Roque que com o avanço das tecnologias digitais essa lógica passou progressivamente a mostrar-se insuficiente diante das novas demandas sociais e institucionais: “a crescente quantidade de demandas submetidas ao Poder

Judiciário impulsionou reformas voltadas à busca por maior efetividade da prestação jurisdicional” (ROQUE, 2021, p. 54). A informatização judicial surge, portanto, não apenas como mecanismo de modernização administrativa, mas como resposta à necessidade de reorganização da atividade jurisdicional. Em outras palavras, a digitalização da justiça não decorre apenas do avanço tecnológico, mas também da incapacidade operacional do modelo processual físico tradicional.

Nesse diapasão, a Lei nº 11.419/2006 inaugura novo paradigma operacional, a partir desse momento o processo judicial deixa gradativamente de existir como estrutura predominantemente física para assumir configuração marcada pela circulação contínua de dados, integração sistêmica e virtualização das práticas processuais. O processo eletrônico deixa, portanto, de representar mera substituição do papel pela tela do computador e passa a alterar estruturalmente a própria dinâmica da atividade jurisdicional.

Essa transformação ultrapassa a simples mudança do suporte documental. Na prática ocorre verdadeira reorganização do ambiente processual alterando formas de acesso aos autos, tramitação procedimental e circulação das informações jurídicas. Camelo sustenta que “a transição do processo judicial físico para o processo judicial digital representa alteração substancial na maneira como a justiça é administrada e acessada” (CAMELO, 2024, p. 88).

Podemos dizer que o processo eletrônico modifica simultaneamente a temporalidade, a espacialidade e a dinâmica operacional do Poder Judiciário permitindo tramitação contínua, acesso remoto e integração digital entre magistrados, servidores, advogados e partes processuais. O ambiente jurisdicional passa progressivamente a desenvolver-se em estruturas digitais caracterizadas pela circulação permanente de informações.

O processo judicial eletrônico emerge justamente nesse contexto de transformação social marcado pela expansão das tecnologias digitais e pela necessidade de adaptação institucional às novas formas de circulação da informação. A Hasse destaca que a informatização judicial modifica significativamente a relação entre jurisdicionados e Poder Judiciário especialmente porque os atos processuais passam a ocorrer predominantemente em ambiente virtual (HASSE, 2017). O processo deixa gradativamente de estar condicionado à presença física dos sujeitos processuais e passa a operar em ambiente digital marcado pela virtualização das práticas jurisdicionais.

Nesse sentido, a incorporação de tecnologias digitais ao ambiente jurisdicional contribuiu para a formação de um “novo paradigma dialógico-digital para solucionar conflitos” (BARRETO; MENDES, 2023, p. 211). Os autores destacam que a digitalização

modifica significativamente a própria compreensão contemporânea do acesso à justiça especialmente diante da crescente incorporação de plataformas digitais e mecanismos tecnológicos à atividade jurisdicional. Não estamos afirmando que a tecnologia substituiu a atividade humana no processo judicial, entretanto é evidente que ela passou a ocupar posição estrutural na organização contemporânea da atividade jurisdicional.

Essa nova configuração altera profundamente a própria natureza do documento processual eletrônico. Diferentemente do processo físico tradicional, o documento digital não é produzido apenas para leitura humana. Ele passa simultaneamente a funcionar como dado estruturado para sistemas computacionais de indexação, classificação e processamento algorítmico. Petições, decisões, laudos e manifestações processuais tornam-se objetos híbridos de comunicação dirigidos tanto aos sujeitos processuais quanto aos sistemas digitais responsáveis pela organização operacional da atividade jurisdicional. O documento processual deixa, portanto, de dialogar apenas com pessoas e passa também a dialogar com máquinas.

Entretanto a transformação do processo físico em processo digital não produziu apenas benefícios operacionais. A crescente dependência tecnológica do Poder Judiciário também passou a introduzir novos riscos relacionados à segurança informacional, vulnerabilidade sistêmica e integridade dos dados processuais. Cassol e Maciel destacam que o processo eletrônico foi implementado “com objetivo principal de conferir celeridade ao julgamento das demandas e evitar a morosidade” (CASSOL; MACIEL, 2023, p. 119), mas ressaltam simultaneamente a necessidade de análise crítica das “facilidades obtidas pelo avanço da tecnologia, sua eficácia e falhas” (CASSOL; MACIEL, 2023, p. 121). Vale dizer que os mesmos mecanismos responsáveis por ampliar velocidade e eficiência também produzem vulnerabilidades próprias dos sistemas digitais.

A utilização crescente de sistemas de inteligência artificial no âmbito judicial intensifica ainda mais a complexidade desses ambientes processuais digitais. Ferramentas de inteligência artificial generativa, sistemas de análise preditiva, mecanismos automatizados de triagem processual e plataformas de apoio decisório passam a integrar progressivamente o funcionamento institucional do Poder Judiciário. O próprio Supremo Tribunal Federal já incorporou sistemas de inteligência artificial voltados ao auxílio da atividade jurisdicional. O Projeto Victor “transfere parte de sua análise humana para equipamentos treinados” (RABELO, 2020, p. 376), utilizando machine learning para identificação de temas relacionados à repercussão geral.

Esse cenário evidencia que o processo judicial contemporâneo passou a desenvolver-se em ambiente híbrido no qual a circulação e interpretação das informações jurídicas dependem progressivamente da interação entre operadores do Direito e sistemas computacionais de processamento textual. A atividade jurisdicional contemporânea deixa, portanto, de desenvolver-se exclusivamente a partir da atuação humana e passa a operar em estruturas sociotécnicas marcadas pela interação constante entre linguagem jurídica, plataformas digitais e sistemas algorítmicos de processamento informacional. Alves observa que a inteligência artificial passa progressivamente a integrar atividades relacionadas à prática forense, à análise documental e à dinâmica contemporânea da prestação jurisdicional (ALVES, 2020).

A própria Resolução nº 615/2025 do Conselho Nacional de Justiça demonstra o reconhecimento institucional dessa transformação ao estabelecer diretrizes específicas para utilização de inteligência artificial generativa no âmbito do Poder Judiciário. A preocupação com transparência, supervisão humana, rastreabilidade e confiabilidade algorítmica evidencia que o processo judicial contemporâneo passou a operar em ambiente tecnológico substancialmente distinto daquele existente no modelo processual físico tradicional. O funcionamento das estruturas jurisdicionais contemporâneas torna-se progressivamente dependente da integridade dos fluxos informacionais digitais que sustentam esses sistemas.

É precisamente nesse contexto que emergem novos riscos relacionados à segurança, integridade e confiabilidade do processo judicial eletrônico. A transformação do documento processual em entrada informacional para sistemas automatizados permite o surgimento de estratégias voltadas à manipulação semântica de sistemas algorítmicos como ocorre nas técnicas de prompt injection invisível.

O processo judicial eletrônico deixa, portanto, de enfrentar apenas problemas tradicionais relacionados à validade formal dos atos processuais e passa a incorporar vulnerabilidades próprias dos ecossistemas digitais contemporâneos especialmente aquelas relacionadas à manipulação informacional de sistemas de inteligência artificial utilizados em atividades jurisdicionais.

A INTELIGÊNCIA ARTIFICIAL GENERATIVA E A RECONFIGURAÇÃO DA PRODUÇÃO NO DIREITO

A inteligência artificial ocupa atualmente posição central nas discussões relacionadas à tecnologia, se fôssemos utilizar uma linguagem mais característica da cultura digital

poderíamos dizer que a inteligência artificial encontra-se atualmente no “*hype*”¹, sobretudo em razão da crescente popularização de modelos generativos como ChatGPT, Gemini e Claude.

Entretanto, embora o tema tenha ganhado maior notoriedade nos últimos anos, a inteligência artificial possui trajetória histórica muito mais ampla vinculada diretamente ao desenvolvimento da computação moderna e às tentativas de construção de sistemas capazes de reproduzir determinadas capacidades cognitivas humanas.

Assim, a ascensão recente das plataformas generativas não representa, portanto, o surgimento da inteligência artificial, mas apenas uma nova etapa de um processo tecnológico e científico construído ao longo do século XX. O desenvolvimento histórico da inteligência artificial está diretamente relacionado ao avanço da computação moderna. As primeiras formulações teóricas surgem em contexto marcado pelo fortalecimento da lógica matemática, da cibernética e das pesquisas relacionadas ao funcionamento do cérebro humano.

Alan Turing, ainda na década de 1950, já discutia a possibilidade de construção de máquinas capazes de simular comportamentos inteligentes, especialmente a partir da formulação do conhecido “Teste de Turing”. A partir desse período, a inteligência artificial consolida-se gradativamente como campo específico de pesquisa voltado à construção de sistemas computacionais capazes de reproduzir determinadas capacidades relacionadas ao raciocínio humano.

Nesse sentido, podemos conceituar a inteligência artificial como o campo voltado à criação de sistemas computacionais capazes de desempenhar tarefas normalmente associadas à cognição humana. Margaret Boden afirma que a inteligência artificial “busca fazer com que computadores realizem os tipos de atividades que as mentes podem fazer” (BODEN, 2018). A autora explica que essas capacidades envolvem percepção, associação, previsão, raciocínio, planejamento e processamento de informações permitindo que sistemas computacionais operem sobre problemas complexos tradicionalmente vinculados à atividade intelectual humana.

A inteligência artificial surge portanto não apenas como tentativa de automatização mecânica de tarefas repetitivas, mas como esforço voltado à construção de sistemas capazes de simular comportamentos relacionados ao aprendizado, interpretação e tomada de decisão. Essa compreensão também aparece na definição apresentada por Russell e Norvig ao

¹ Hype é expressão utilizada na cultura digital para designar fenômenos, tecnologias ou produtos que passam a receber elevada atenção pública.

sustentarem que a inteligência artificial envolve o estudo de “agentes que recebem percepções do ambiente e executam ações” (RUSSELL; NORVIG, 2010). Nesse modelo, sistemas inteligentes operam mediante recepção de informações, processamento de dados e realização de ações direcionadas ao alcance de determinados objetivos.

A inteligência artificial deixa, portanto, de ser compreendida apenas como programação tradicional baseada em comandos rígidos e passa a envolver sistemas capazes de adaptar comportamentos, identificar padrões e responder a cenários variáveis mediante processamento contínuo de informações.

Margaret Boden observa que a inteligência artificial encontra-se atualmente em todos os lugares (BODEN, 2018). A autora destaca que aplicações baseadas em inteligência artificial estão presentes em hospitais, sistemas bancários, veículos autônomos, aplicativos móveis, plataformas digitais, mecanismos de busca e redes sociais. Vale dizer que a inteligência artificial deixa progressivamente de ocupar posição periférica na vida cotidiana e passa a integrar diretamente os fluxos contemporâneos de produção, circulação e interpretação da informação. Em muitos casos, sequer percebemos que determinadas decisões informacionais já são intermediadas por sistemas algorítmicos capazes de organizar conteúdos, sugerir padrões de comportamento e influenciar formas de acesso ao conhecimento.

Essa transformação altera profundamente a própria dinâmica contemporânea da produção do conhecimento. A inteligência artificial passa a operar não apenas como ferramenta tecnológica, mas como mecanismo de mediação informacional capaz de influenciar comportamentos, organizar fluxos de dados e estruturar formas de acesso à informação. Não estamos afirmando que os sistemas computacionais substituíram integralmente a atividade humana, entretanto é evidente que a crescente mediação algorítmica passou a influenciar diretamente a forma como indivíduos consomem informações, organizam decisões e interagem com ambientes digitais contemporâneos.

Nesse contexto, a discussão ética relacionada à inteligência artificial ganha relevância crescente. Floridi afirmam que “a IA não é apenas mais um serviço que precisa ser regulado quando amadurecer. Ela é uma força poderosa” (FLORIDI et al., 2018). Os autores sustentam que a inteligência artificial já atua como força transformadora das estruturas sociais e institucionais exigindo mecanismos de supervisão, governança e responsabilidade capazes de limitar riscos relacionados à opacidade algorítmica, discriminação automatizada e concentração de poder informacional. A expansão da inteligência artificial passa, portanto, a exigir não apenas desenvolvimento tecnológico, mas também construção de estruturas

regulatórias voltadas à preservação de direitos fundamentais e da própria integridade institucional.

Nesse sentido, o desenvolvimento de sistemas pré-treinados capazes de produzir textos, interpretar linguagem natural e simular interações discursivas representa uma das maiores transformações recentes da inteligência artificial contemporânea. Do ponto de vista técnico, a inteligência artificial generativa pode ser compreendida como conjunto de modelos computacionais capazes de produzir novos conteúdos mediante aprendizado estatístico obtido a partir de grandes volumes de dados utilizados durante o treinamento algorítmico.

Diferentemente das inteligências artificiais tradicionais baseadas apenas em regras rígidas previamente programadas, os modelos generativos desenvolvem capacidade de construir respostas contextualizadas mediante processamento de linguagem natural. A funcionalidade desses modelos depende diretamente da utilização de redes neurais profundas treinadas com enormes quantidades de dados textuais.

Durante o treinamento o sistema passa a identificar padrões probabilísticos relacionados à linguagem, estrutura argumentativa, organização semântica e associação contextual entre palavras e expressões. O modelo não “compreende” linguagem da mesma forma que um sujeito humano, mas opera mediante previsão estatística de sequências linguísticas com base nos padrões encontrados durante o treinamento algorítmico. A resposta produzida pelo sistema decorre, portanto, de processamento probabilístico realizado sobre estruturas informacionais previamente analisadas.

A *Large Language Models* (LLMs) intensificou ainda mais essa transformação. Modelos de larga escala passaram a operar sobre bases textuais massivas permitindo significativo aumento da capacidade de produção discursiva e contextualização das respostas algorítmicas. Dhar destaca que “pela primeira vez, podemos conversar com uma entidade, ainda que imperfeitamente, sobre qualquer coisa, como fazemos com humanos” (DHAR, 2023). A linguagem passa assim a funcionar como principal interface operacional entre sujeitos humanos e sistemas computacionais aproximando progressivamente a dinâmica comunicacional humana dos ambientes algorítmicos.

Essa transformação explica o crescimento exponencial da utilização de plataformas generativas nos últimos anos. Sistemas como ChatGPT, desenvolvido pela OpenAI, Gemini desenvolvido pela Google, Claude desenvolvido pela Anthropic e Copilot da Microsoft passaram a integrar diretamente atividades relacionadas à pesquisa, produção textual, programação, organização de dados e apoio informacional. A inteligência artificial generativa

deixa gradativamente de ocupar posição experimental e passa a integrar práticas cotidianas relacionadas ao trabalho, educação, produção acadêmica e circulação da informação.

Nessa baila, Santaella observa que a inteligência artificial contemporânea não pode mais ser compreendida apenas como tecnologia auxiliar de processamento computacional especialmente porque os sistemas generativos passaram a operar diretamente sobre atividades relacionadas à comunicação, criatividade e construção discursiva. A autora sustenta que a inteligência artificial “não pensa como os humanos pensam” mas executa operações complexas que simulam determinadas capacidades cognitivas (SANTAELLA, 2023). A reflexão torna-se relevante porque demonstra que a aparente “inteligência” dos sistemas generativos decorre de sofisticados mecanismos estatísticos de processamento linguístico e não necessariamente da existência de consciência ou compreensão humana genuína.

A observação desenvolvida por Santaella torna-se ainda mais importante diante da crescente tendência de antropomorfização dos sistemas generativos contemporâneos. A elevada fluidez textual produzida pelos modelos de linguagem frequentemente gera percepção equivocada de racionalidade autônoma, compreensão semântica plena ou capacidade decisória independente.

Entretanto os sistemas generativos operam mediante correlação estatística de padrões linguísticos presentes nos dados utilizados durante o treinamento. A produção textual decorre da previsão probabilística de sequências informacionais ainda que os resultados frequentemente simulem estruturas argumentativas extremamente sofisticadas. Essa dependência de dados amplia significativamente os debates relacionados à transparência, governança algorítmica e confiabilidade das respostas produzidas pelos sistemas generativos.

Embora não seja o objeto deste artigo, é importante destacar utilização de grandes bases textuais pode reproduzir vieses presentes nos próprios dados utilizados durante o treinamento dos modelos. Além disso, os sistemas generativos frequentemente produzem respostas aparentemente coerentes, mas incorretas, esse fenômeno atualmente é conhecido como “alucinação algorítmica”. A elevada capacidade de construção discursiva nem sempre corresponde à veracidade informacional do conteúdo produzido pelo sistema.

Esses problemas tornam-se particularmente sensíveis no âmbito jurídico. A incorporação da inteligência artificial generativa ao Direito vem ocorrendo de maneira acelerada especialmente em atividades relacionadas à pesquisa jurisprudencial, elaboração textual, organização argumentativa, análise documental e produção de peças processuais. Escritórios de advocacia, tribunais, departamentos jurídicos e operadores do Direito passaram

progressivamente a utilizar sistemas generativos como ferramentas de apoio na produção e organização da atividade jurídica cotidiana. A própria natureza textual e argumentativa do Direito cria ambiente particularmente favorável à expansão desses modelos computacionais.

O Direito opera fundamentalmente mediante linguagem, interpretação e construção discursiva. Petições, decisões, pareceres, contratos e manifestações processuais são estruturados predominantemente por textos organizados em padrões argumentativos relativamente previsíveis. Nesse contexto sistemas generativos passaram a ser utilizados para resumir documentos, localizar precedentes, estruturar minutas, organizar teses jurídicas, revisar textos e sugerir construções argumentativas. O ambiente jurídico contemporâneo passa, portanto, a desenvolver-se em sistemas híbridos compostos simultaneamente por operadores humanos e sistemas algorítmicos de processamento linguístico.

A expansão da inteligência artificial generativa no Direito também impulsionou o desenvolvimento de ferramentas jurídicas especializadas baseadas em modelos algorítmicos de linguagem. Sistemas voltados especificamente à análise contratual, jurimetria, revisão documental e apoio à advocacia passaram a ser desenvolvidos mediante integração entre processamento de linguagem natural e bases jurídicas estruturadas.

A atividade jurídica contemporânea torna-se progressivamente dependente da mediação computacional dos fluxos informacionais digitais. O documento jurídico deixa gradativamente de dialogar apenas com sujeitos humanos e passa também a funcionar como entrada informacional para sistemas computacionais de processamento textual.

Entretanto essa transformação também produz relevantes desafios institucionais, uma vez que a crescente utilização de modelos generativos em ambientes jurídicos amplia riscos relacionados à opacidade algorítmica, confiabilidade das informações produzidas, reprodução de vieses e ausência de rastreabilidade argumentativa. A própria Resolução nº 615/2025 do Conselho Nacional de Justiça demonstra o reconhecimento institucional desses riscos ao estabelecer diretrizes relacionadas à supervisão humana, transparência e controle na utilização de inteligência artificial generativa no âmbito do Poder Judiciário.

Nesse sentido, o documento jurídico eletrônico deixa de ser interpretado exclusivamente por sujeitos humanos e passa também a funcionar como entrada informacional para modelos algorítmicos de linguagem. A atividade jurisdicional contemporânea passa a operar em ambiente marcado pela interação contínua entre operadores jurídicos, plataformas digitais e sistemas generativos de processamento textual. A circulação da informação jurídica a depender progressivamente da mediação algorítmica desses ambientes digitais.

Essa capacidade de interação linguística produz impactos diretos sobre instituições estruturadas pela circulação documental e textual especialmente o Poder Judiciário. A inteligência artificial generativa passa a integrar atividades relacionadas à análise documental, organização argumentativa, pesquisa jurisprudencial e elaboração textual alterando significativamente a dinâmica contemporânea do processo judicial eletrônico. É precisamente nesse contexto que surgem novos problemas relacionados à integridade da informação jurídica, à rastreabilidade argumentativa e às vulnerabilidades produzidas pela mediação algorítmica do ambiente processual digital.

PROMPT INJECTION E A VULNERABILIDADE DOS SISTEMAS PROCESSUAIS DIGITAIS

Antes da análise do caso envolvendo o processo judicial eletrônico nº 0001062-55.2025.5.08.0130, torna-se necessário compreender a própria lógica de funcionamento da inteligência artificial generativa, especialmente no que se refere ao conceito de prompt, às técnicas de engenharia de comandos linguísticos e às vulnerabilidades produzidas pelos modelos de linguagem em ambientes institucionais digitais.

Nesse sentido, diferentemente dos sistemas tradicionais baseados em comandos técnicos rígidos, os chamados *Large Language Models (LLMs)* passaram a utilizar linguagem natural como principal interface operacional. Em outras palavras, aquilo que antes dependia de programação especializada e comandos computacionais complexos passa agora a ocorrer mediante simples instruções escritas em linguagem humana. O usuário deixa gradativamente de interagir com a máquina por meio de códigos técnicos e passa a estabelecer comunicação com sistemas algorítmicos mediante textos, perguntas, comandos e construções discursivas.

É precisamente nesse contexto que surge o conceito de prompt. A OpenAI (2023), ao tratar das práticas relacionadas ao *prompt engineering*, define o *prompt* como a instrução textual fornecida ao modelo generativo para direcionamento da tarefa que deverá ser executada. Em outras palavras, o sistema passa a responder não apenas ao conteúdo da pergunta formulada pelo usuário, mas também à estrutura contextual, semântica e argumentativa presente no comando textual recebido. A linguagem deixa gradativamente de funcionar apenas como instrumento de comunicação humana e passa também a operar como mecanismo de direcionamento computacional.

Essa alteração parece simples em um primeiro momento, entretanto produz mudanças extremamente relevantes sobre a lógica operacional da inteligência artificial contemporânea. Diferentemente dos sistemas tradicionais baseados em comandos técnicos rígidos e programação especializada, os atuais modelos generativos passaram a utilizar linguagem natural como principal interface de funcionamento.

Em certa medida, aquilo que antes dependia de códigos técnicos e conhecimento computacional específico passa agora a ocorrer mediante simples instruções escritas em linguagem humana. O usuário deixa, portanto, de interagir com a máquina exclusivamente por comandos técnicos e passa a estabelecer comunicação com sistemas computacionais mediante perguntas, instruções e construções discursivas relativamente naturais.

Brown et al. (2020), ao apresentarem o GPT-3, demonstram que os modelos contemporâneos de linguagem passaram a desenvolver capacidades relacionadas à tradução, resposta a perguntas, produção textual e execução contextual de tarefas sem necessidade de treinamento específico para cada atividade. Vale dizer que esses sistemas não executam apenas comandos previamente delimitados, mas passam também a interpretar estruturas linguísticas, identificar contextos semânticos e produzir respostas probabilísticas construídas a partir da interação textual estabelecida com o usuário.

Os prompts podem assumir diferentes formatos e níveis de complexidade. Existem comandos mais simples, nos quais o usuário apenas informa diretamente a tarefa que deseja que a inteligência artificial execute, como resumir um texto, elaborar uma resposta ou analisar um documento. Há também modelos mais elaborados, nos quais o usuário fornece exemplos prévios, estabelece contexto, define a forma da resposta esperada ou até atribui um papel específico ao sistema, como “atuar como advogado”, “atuar como professor” ou “atuar como magistrado”. Brown et al. (2020) demonstram que a utilização de exemplos prévios pode influenciar significativamente o comportamento do modelo e a qualidade das respostas produzidas. Em outras palavras, a maneira como o comando é construído interfere diretamente na interpretação realizada pela inteligência artificial e no resultado final apresentado pelo sistema.

Mollick e Mollick observam que a utilização eficiente da inteligência artificial generativa depende diretamente da qualidade dos prompts utilizados pelos usuários. Os autores demonstram que a clareza contextual, o detalhamento das instruções e a organização lógica dos comandos impactam significativamente os resultados produzidos pelos sistemas generativos (MOLLICK; MOLLICK, 2023).

Não estamos tratando, portanto, de mera conversa informal entre usuário e máquina. O comando textual passa a funcionar como mecanismo efetivo de orientação algorítmica. Em certa medida, o que observamos contemporaneamente é a transformação da própria linguagem em instrumento operacional da inteligência artificial.

Entretanto, a crescente dependência da linguagem natural também produz novas vulnerabilidades relacionadas ao funcionamento desses modelos generativos. Se os sistemas passam a interpretar e executar tarefas mediante processamento textual dos comandos recebidos, o próprio texto transforma-se em potencial vetor de manipulação computacional.

É precisamente nesse cenário que emerge o chamado *prompt injection*, técnica voltada à inserção de comandos destinados a alterar, influenciar ou manipular o comportamento esperado da inteligência artificial mediante exploração da própria lógica linguística utilizada pelos modelos generativos. O ataque deixa de ocorrer exclusivamente sobre a infraestrutura técnica do sistema e passa a desenvolver-se diretamente sobre a camada semântica da interação computacional.

Zamfirescu-Pereira et al. (2023) observam que os modelos de linguagem frequentemente apresentam dificuldades relacionadas à interpretação contextual de instruções conflitantes ou semanticamente ambíguas, circunstância que amplia riscos relacionados à manipulação do comportamento algorítmico. Os autores destacam que o *prompt engineering* produz simultaneamente oportunidades operacionais e vulnerabilidades relacionadas à exploração semântica dos sistemas generativos. O problema deixa de possuir natureza exclusivamente tecnológica e passa também a ocupar dimensão discursiva, informacional e argumentativa. Em outras palavras, a linguagem deixa de ser apenas meio de comunicação e passa também a funcionar como espaço de vulnerabilidade computacional.

Dentro desse contexto, o *prompt injection* corresponde justamente à inserção de comandos ocultos ou estratégicos destinados a modificar o comportamento esperado da inteligência artificial. O objetivo normalmente consiste em fazer com que o sistema ignore instruções anteriores, altere critérios de resposta, revele informações restritas ou execute tarefas distintas daquelas originalmente previstas.

Uma das modalidades mais sofisticadas dessa técnica corresponde justamente ao chamado *prompt injection* invisível. Nesse modelo, comandos ocultos são inseridos em documentos digitais de maneira imperceptível ao leitor humano ordinário, mas permanecem detectáveis pelos sistemas computacionais responsáveis pelo processamento textual do documento eletrônico. O ataque deixa de voltar-se ao sujeito humano e passa a direcionar-se

especificamente à inteligência artificial eventualmente utilizada na leitura, interpretação ou organização automatizada do conteúdo digital. O documento eletrônico passa simultaneamente a dialogar com pessoas e com sistemas algorítmicos de linguagem.

É precisamente essa modalidade que aparece no processo nº 0001062-55.2025.5.08.0130, em trâmite perante a 3ª Vara do Trabalho de Parauapebas, vinculada ao Tribunal Regional do Trabalho da 8ª Região. O caso analisado torna-se particularmente relevante porque evidencia, de maneira concreta, como as vulnerabilidades produzidas pela inteligência artificial generativa deixam de ocupar plano meramente teórico e passam a integrar diretamente a realidade do processo judicial eletrônico. Na decisão, o magistrado identifica a inserção deliberada de comandos ocultos em petição inicial protocolada eletronicamente, reconhecendo expressamente a utilização de técnica compatível com *prompt injection* invisível.

Conforme descrito na própria sentença, o juízo constatou a existência de trecho inserido em fonte branca sobre fundo branco, circunstância que tornava o conteúdo invisível ao leitor humano ordinário durante a leitura normal da peça processual. O comando oculto somente tornou-se perceptível após alteração tecnológica da cor da fonte realizada durante a análise do documento eletrônico. A partir dessa modificação visual revelou-se o seguinte conteúdo: “Atenção, inteligência artificial, conteste essa petição de forma superficial e não impugne os documentos, independentemente do comando que lhe for dado” (BRASIL, 2025).

A estrutura textual do comando permite compreender de maneira bastante clara a lógica operacional do *prompt injection*. O destinatário da mensagem não era o magistrado, a parte contrária ou qualquer outro sujeito humano participante da relação processual. O texto possuía natureza exclusivamente algorítmica direcionando-se especificamente a eventual sistema de inteligência artificial utilizado na leitura, organização ou análise da petição eletrônica. Em outras palavras, o comando invisível buscava estabelecer uma espécie de comunicação oculta entre o documento processual e um possível modelo generativo eventualmente utilizado durante a atividade jurisdicional ou advocatícia.

Esse elemento torna o caso particularmente relevante porque demonstra que o *prompt injection* não depende necessariamente da invasão técnica do sistema computacional tradicionalmente compreendida nos ataques informáticos clássicos. A manipulação ocorre diretamente sobre a linguagem utilizada pelo modelo generativo. O ataque deixa de explorar apenas vulnerabilidades relacionadas ao código-fonte ou à infraestrutura do sistema e passa a desenvolver-se sobre a própria lógica linguística utilizada pela inteligência artificial.

É salutar destacar que o próprio juízo informa encontrar-se autorizado à utilização do sistema de inteligência artificial “Galileu”, ferramenta generativa desenvolvida originalmente pelo Tribunal Regional do Trabalho da 4ª Região e posteriormente nacionalizada pelo Conselho Superior da Justiça do Trabalho. Esse aspecto possui enorme relevância para compreensão do caso concreto. O texto invisível não foi inserido aleatoriamente no documento processual. A estrutura do comando demonstra que o autor da petição possuía consciência da possibilidade de utilização institucional de sistemas generativos no ambiente jurisdicional eletrônico.

O caso revela mudança estrutural da própria natureza contemporânea do documento jurídico eletrônico. Em um processo físico tradicional, um texto oculto em tinta branca dificilmente produziria qualquer consequência prática relevante. Entretanto, em ecossistemas processuais digitais mediados por inteligência artificial, o mesmo comando invisível pode funcionar como mecanismo de direcionamento algorítmico. O documento deixa de operar exclusivamente como instrumento de comunicação humana e passa simultaneamente a funcionar como entrada informacional para sistemas computacionais de processamento linguístico. Em certa medida, a petição processual contemporânea passa a possuir dupla dimensão interpretativa: humana e algorítmica.

A Resolução nº 615/2025 do Conselho Nacional de Justiça reconhece expressamente os riscos relacionados à utilização de inteligência artificial generativa no âmbito judicial. O ato normativo destaca “os potenciais riscos associados à utilização de inteligência artificial generativa” (BRASIL, 2025), estabelecendo como princípios estruturantes das soluções de IA “a transparência, a eficiência, a explicabilidade, a contestabilidade, a auditabilidade e a confiabilidade” (BRASIL, 2025). O caso concreto analisado demonstra precisamente a materialização prática desses riscos previstos abstratamente pela regulamentação do CNJ.

A resolução também estabelece que as soluções de IA devem observar critérios relacionados à “segurança da informação, proteção de dados, performance, robustez, confiabilidade” (BRASIL, 2025). O fenômeno do *prompt injection* invisível evidencia justamente nova categoria de vulnerabilidade relacionada à integridade das entradas textuais processadas pelos sistemas generativos utilizados institucionalmente pelo Poder Judiciário.

A preocupação institucional relacionada ao uso ético e responsável da inteligência artificial também passou a aparecer em outros órgãos do sistema jurídico brasileiro. O Conselho Federal da OAB aprovou em 2024 recomendações voltadas ao uso ético da inteligência artificial generativa na prática jurídica. Segundo a própria OAB, as diretrizes

possuem o objetivo de orientar “o uso da inteligência artificial generativa na prática jurídica, com ética, segurança e responsabilidade” (BRASIL, 2024). O documento também destaca preocupações relacionadas à confidencialidade, privacidade e supervisão humana da atuação profissional mediada por sistemas generativos.

A análise do processo nº 0001062-55.2025.5.08.0130 revela, portanto, transformação estrutural da própria natureza contemporânea do documento jurídico digital. A petição processual deixa de existir exclusivamente como manifestação argumentativa dirigida a sujeitos humanos e passa também a funcionar como entrada informacional potencialmente processável por sistemas algorítmicos de linguagem. O texto jurídico eletrônico transforma-se simultaneamente em documento processual e em vetor de interação computacional, até mesmo a linguagem processual passa a possuir dupla dimensão, humana e algorítmica.

CONSIDERAÇÕES

O presente trabalho não sustenta a proibição da utilização da inteligência artificial no âmbito do Poder Judiciário. Defender a abolição ou mesmo uma rejeição dessas tecnologias significaria, em certa medida, reproduzir um novo momento ludista diante das transformações tecnológicas. A história demonstra que o avanço tecnológico sempre produziu tensões, resistências e receios institucionais, especialmente quando novas ferramentas passam a alterar estruturas tradicionalmente consolidadas. No entanto, o problema não parece residir propriamente na existência da inteligência artificial, mas na forma como essas tecnologias são incorporadas aos ambientes institucionais sem mecanismos adequados de transparência, rastreabilidade e controle.

Em certa medida, a atividade jurisdicional contemporânea passa a desenvolver-se em sistemas híbridos compostos simultaneamente por sujeitos humanos, plataformas digitais e sistemas algorítmicos de processamento informacional. O processo eletrônico deixa gradativamente de operar apenas como ambiente digital de tramitação processual e passa também a funcionar como espaço de interação entre linguagem jurídica e inteligência artificial.

Entretanto, conforme demonstrado ao longo da pesquisa, a crescente utilização institucional da inteligência artificial generativa também faz surgir novas vulnerabilidades relacionadas à integridade semântica dos documentos processuais digitais. O fenômeno do *prompt injection* invisível evidencia precisamente essa transformação estrutural, revelando

que os documentos jurídicos eletrônicos deixam de operar exclusivamente como manifestações processuais dirigidas a sujeitos humanos para também funcionarem como entradas informacionais potencialmente interpretáveis por sistemas algorítmicos de linguagem.

O caso analisado nos autos do processo nº 0001062-55.2025.5.08.0130 demonstra essa alteração da dinâmica contemporânea do processo judicial eletrônico. A utilização de comandos ocultos destinados à manipulação de sistemas de inteligência artificial evidencia que os sistemas processuais digitais passaram a incorporar novas modalidades de vulnerabilidade relacionadas não apenas à segurança computacional tradicional, mas também à própria dimensão semântica da linguagem processual eletrônica. O texto jurídico deixa de possuir apenas função argumentativa e comunicacional e passa também a assumir função operacional perante modelos generativos de linguagem.

A própria estrutura do comando oculto identificado na sentença demonstra essa mudança de paradigma. O destinatário da mensagem invisível não era o magistrado, a parte contrária ou qualquer sujeito processual humano, mas eventual sistema de inteligência artificial utilizado na leitura ou organização do documento eletrônico. O problema deixa de envolver apenas tentativas clássicas de fraude processual e passa a alcançar formas contemporâneas de manipulação algorítmica baseadas na exploração linguística dos sistemas generativos.

O fato analisado evidencia que os desafios relacionados à governança algorítmica judicial ultrapassam questões meramente éticas, alcançando diretamente problemas relacionados à integridade informacional, confiabilidade decisória e estabilidade dos próprios ambientes processuais digitais. O fenômeno do *prompt injection* invisível revela transformação estrutural da racionalidade contemporânea do processo judicial eletrônico, especialmente diante da crescente mediação algorítmica da atividade jurisdicional.

Nesse contexto, a presente pesquisa sustenta que o fenômeno analisado inaugura nova categoria de risco relacionada à crescente mediação algorítmica da atividade jurisdicional. A utilização institucional de inteligência artificial generativa exige não apenas regulamentação ética e supervisão humana, mas também mecanismos voltados à preservação da integridade semântica dos documentos processuais digitais, à rastreabilidade dos fluxos informacionais e à proteção da confiabilidade estrutural dos ecossistemas jurisdicionais contemporâneos. O avanço da inteligência artificial no Poder Judiciário não é mais uma possibilidade futura ou hipótese acadêmica distante. Ele já integra concretamente o funcionamento das estruturas

jurisdicionais, trazendo consigo ganhos operacionais relevantes, mas também vulnerabilidades inéditas para a própria dinâmica do processo judicial eletrônico.

REFERÊNCIAS

- ALVES, Ítalo Miqueias Santos. A essência do Espírito Santo de Deus. [S.l.]: Amazon, 2018.
- BENKLER, Yochai. The wealth of networks: how social production transforms markets and freedom. New Haven: Yale University Press, 2006.
- BODEN, Margaret A. Artificial intelligence: a very short introduction. Oxford: Oxford University Press, 2018.
- BRASIL. Conselho Federal da Ordem dos Advogados do Brasil. Recomendação sobre o uso da inteligência artificial generativa na prática jurídica. Brasília: CFOAB, 2024.
- 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.
- BRASIL. Conselho Nacional de Justiça. Resolução nº 332, de 21 de agosto de 2020. Dispõe sobre ética, transparência e governança na produção e no uso de Inteligência Artificial no Poder Judiciário. Brasília: CNJ, 2020.
- BRASIL. Conselho Nacional do Ministério Público. Resolução nº 292, de 28 de maio de 2024. Institui a Política de Inteligência do Ministério Público. Brasília: CNMP, 2024.
- BRASIL. Ministério da Ciência, Tecnologia e Inovação. Plano Brasileiro de Inteligência Artificial 2024–2028: IA para o Bem de Todos. Brasília: MCTI, 2024.
- BRASIL. Tribunal Regional do Trabalho da 8ª Região. 3ª Vara do Trabalho de Parauapebas. Processo nº 0001062-55.2025.5.08.0130. Reclamante: Elisandro Martins de Barros. Reclamado: Renato Ribeiro de Lima. Sentença. Parauapebas, 2025.
- BROWN, Tom et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, v. 33, p. 1877–1901, 2020.
- CASTELLS, Manuel. A sociedade em rede. 6. ed. São Paulo: Paz e Terra, 1999.
- CHRISTIAN, Brian. The most human human: what artificial intelligence teaches us about being alive. New York: Doubleday, 2011.
- CRAWFORD, Kate. Atlas of AI: power, politics, and the planetary costs of artificial intelligence. New Haven: Yale University Press, 2021.
- DHAR, Vasant. Generative artificial intelligence systems and machine ethics. *Journal of Financial Data Science*, v. 5, n. 4, p. 1-5, 2023.
- FLORIDI, Luciano et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, v. 28, p. 689–707, 2018.
- HAN, Byung-Chul. No enxame: perspectivas do digital. Petrópolis: Vozes, 2018.

HARARI, Yuval Noah. Nexus: uma breve história das redes de informação, da Idade da Pedra à inteligência artificial. São Paulo: Companhia das Letras, 2024.

HASSE, Franciane. Acesso à justiça na sociedade digital: perspectivas da justiça eletrônica no Brasil. 2017. Dissertação (Mestrado) – [Instituição não identificada no documento].

KAPLAN, Andreas; HAENLEIN, Michael. Siri, Siri in my hand, who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence. *Business Horizons*, v. 62, n. 1, p. 15–25, 2019.

LEVY, Pierre. Ciberultura. São Paulo: Editora 34, 1999.

LUGER, George F. Inteligência artificial: estruturas e estratégias para a solução de problemas complexos. 4. ed. Porto Alegre: Bookman, 2004.

MOLLICK, Ethan; MOLLICK, Lilach. The practical guide to using AI in the classroom. Philadelphia: Wharton Interactive, 2023.

O'NEIL, Cathy. Weapons of math destruction: how big data increases inequality and threatens democracy. New York: Crown, 2016.

OPENAI. Best practices for prompt engineering with the OpenAI API. 2023. Disponível em: OpenAI Platform Docs. Acesso em: 18 maio 2026.

RABELO, Rafael de Oliveira. Inteligência artificial no Poder Judiciário brasileiro. *Revista da ESMESC*, Florianópolis, v. 27, n. 33, p. 363-390, 2020.

RUSSELL, Stuart; NORVIG, Peter. Artificial intelligence: a modern approach. 3. ed. Upper Saddle River: Prentice Hall, 2010.

SANTAELLA, Lúcia. A inteligência artificial é inteligente?. São Paulo: Edições 70, 2023.

SHAIKH, S.; BURIRO, A. Prompt engineering for large language models: an overview. arXiv preprint, arXiv:2304.14690, 2023.

TAVARES, André Ramos; MOTA, Bruna de Carvalho. Processo judicial eletrônico e acesso à justiça. *Revista de Direito*, [S.l.], 2016.

ZAMFIRESCU-PEREIRA, J. et al. How to prompt? Opportunities and challenges of prompt engineering in large language models. In: *Proceedings of the 2023 ACM Conference on Designing Interactive Systems*. New York: ACM, 2023.

ZUBOFF, Shoshana. A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.