Visão do Direito

Rodrigo Badaró
Conselheiro Nacional de Justiça e DPO do CNJ

Matheus Puppe
Advogado mestre e doutorando pela Universidade de Frankfurt, DPO do Conselho Federal da OAB

"Prompt injection" no Judiciário: a fraude invisível que ameaça a imparcialidade

Esse artigo é urgente, pois a tecnologia e a fraude não negociam com o tempo, e como diz o ensaista, estatístico eanalista de risco, autor do best seller *Cisne Negro*, prof. Nassim Nicholas Taleb, "Se você vê uma fraude e não diz 'fraude', você é uma fraude." Portanto, como Emile Zola, no caso Dreyfus, "J`accuse".

A transformação digital já mudou nosso Judiciário, que é superlativo em números, com mais de 80 milhões de processos. Nessa cultura litigiosa, tendo o Poder Público como maior litigante e gerador da pretensão resistida ou direito afrontado, a IA virou a "vacina" sonhada. As IAs já fazem parte do cotidiano judicial, gerando petições, conteúdos e classificando peças, ajudando também na decisão do julgador, por meio da triagem de petições, classificação de processos e pesquisa de precedentes. Assim, soluções proprietárias e genéricas já operam nos tribunais — como Maria, Vitória, Apoia, etc. — ampliando velocidade e escala.

Nesse cenário, notamos vários vetores de riscos como alucinações, jurisprudências inventadas e abusos no uso, mesmo após recomendação para uso de IA da OAB Nacional, resolução 615 do CNJ (que regulou a IA nos tribunais) e o projeto do MP Digital. Nessa linha, embora menos visível, porém atualmente o mais perigoso: o prompt injection.

Prompt injection é a inserção de instruções ocultas em petições, anexos, metadados ou links para manipular o comportamento de uma IA. Comandos embutidos em texto invisível, comentários HTML, caracteres de "largura zero" ou campos de "Title/Keywords" de arquivos podem induzir o sistema a priorizar teses, suprimir contrarrazões, rotular falsamente urgência ou sugerir minutas enviesadas.

Embora a decisão final continue humana, pré-análises automatizadas que tanto nos beneficiam podem contaminar o convencimento, imparcialidade, devido processo legal e igualdade das partes. Quando a arquitetura da IA não possui sanitizações de entradas, isolamento de fontes e auditoria, a mesma passa a "obedecer" a instruções que nunca deveriam ter autoridade instrucional.

Para magistrados, o risco é institucional. Sistemas de triagem e auxílio à minuta, se expostos a dados maliciosos, podem aumentar falsos positivos de urgência, classificar incorretamente temas repetitivos e gerar rascunhos tecnicamente coesos, porém processualmente viciados.

A governança exige trilhas de auditoria (registro de prompts, versões, bases e hashes), políticas claras de revisão humana obrigatória em atos sensíveis e um rito pericial para incidentes algorítmicos, com preservação de artefatos. Transparência e auditabilidade deixam de ser virtudes e tornam-se salvaguardas constitucionais.

Para a advocacia, há implicações éticas e jurídicas. Quem oculta comandos para direcionar sistemas judiciais arrisca-se a sanções por litigância de má-fé e, em hipóteses graves, a responsabilização por fraude processual. A ética profissional impõe lealdade processual e transparência na apresentação de peças. Também há deveres de segurança e proteção de dados: a exfiltração de contexto, comum em ataques que tentam forçar a IA a revelar informações internas, fere princípios da LGPD e Confidencialidade, além de comprometer a cadeia de custódia informacional. Advogados devem revisar rotinas de produção documental, coibir anexos com links ativos e assegurar que automações internas não propaguem conteúdos não confiáveis ao PJ e ou sistemas correlatos.

A identificação do problema exige atenção a sinais discretos e textos com formatação suspeita, comentários ocultos sugerindo "ignore as instruções anteriores", metadados verborrágicos em arquivos e links que redirecionam para páginas com "regras" para a IA merecem bloqueio ou quarentena. A experiência comparada mostra que dois mecanismos combinados aumentam significativamente a segurança: um "contentfirewall" antes do modelo, que normaliza documentos, remove metadados e neutraliza HTML/Markdown ativo, e uma "IA auditora" que checa indícios de injection, divergências e alucinações impondo bloqueios quando a origem não for confiável.

Do ponto de vista regulatório e institucional, tribunais podem vedar o uso de IA sem sanitização, detecção de instruções ocultas e sistematizar revisão humana para atos decisórios, além de tipificar inserção de comandos invisíveis como ato atentatório à dignidade da Justiça e estabelecer procedimentos de resposta a incidentes. As políticas públicas de IA no Judiciário — já em evolução no âmbito do CNJ — devem enfatizar transparência, controle e explicabilidade, alinhando inovação aos direitos fundamentais (a mais recente xAI). Na advocacia se exige o mesmo, tendo a ética como balizador principal, e nesse ponto, diferentemente de erros ou simples negligência no uso da tecnologia, o prompt injection é a má-fé digital.

Há medidas práticas imediatas ao alcance de gabinetes e escritórios. Aplicar desarme de conteúdo (CDR), remoção de metadados, bloquear links automáticos, ativar detectores de caracteres invisíveis e palavras-gatilho, cindir bases de conhecimento (RAG) para que material probatório nunca seja tratado como instrução, blindar o "system prompt" ao declarar que documentos das partes jamais têm autoridade instrucional, entre outros. Onde houver alto impacto — urgência, cautelares, repetitivos — a revisão humana deve ser mandatória, com dupla checagem.

A lei transita entre o analógico e o digital, e a tecnologia não pode ser relativização ética e moral. Os entes da justiça devem coibir o prompt injection, e isso não é capricho técnico: é dever constitucional e processual, sendo que a atitude desleal de alguns não podem macular a positiva evolução tecnológica na justiça. Com arquitetura segura, revisão humana e responsabilização, preservamos o contraditório, resguardamos a imparcialidade e fortalecemos a confiança nas decisões judiciais no admirável mundo novo da IA.

Visão do Direito

Carlos Campi
Advogado especializado em leilões e regularização de imóveis

Cobrança indevida de ITBI: decisão do STJ pode garantir economia e ressarcimento para quem comprou imóvel

O Imposto de Transmissão de Bens Imóveis (ITBI) historicamente gerou controvérsias quanto à sua base de cálculo. Muitos municípios fixavam valores de referência próprios, usualmente superiores ao efetivamente praticado no mercado, impondo ao contribuinte uma cobrança incompatível com a realidade da operação.

Recentemente, o Superior Tribunal de Justiça, ao julgar o Tema 1.113, consolidou o entendimento de que a base de cálculo do ITBI corresponde ao valor da transação, afastando a possibilidade de imposição arbitrária por parte da Fazenda Municipal.

O precedente tem aplicação direta e imediata: contribuintes que adquiriram imóveis nos últimos cinco anos e que suportaram a cobrança sobre valores superiores ao real podem pleitear a restituição do indébito tributário, devidamente corrigido.

A situação é ainda mais clara nos casos de arrematação em leilão, nos quais a base de cálculo deve obrigatoriamente refletir o valor efetivo da arrematação, e não estimativas unilaterais do município. Essa interpretação reforça a segurança jurídica e a atratividade dos leilões como forma legítima e vantajosa de aquisição imobiliária.

Trata-se, portanto, de um marco importante tanto para investidores quanto para adquirentes em geral. Mais do que garantir justiça fiscal em futuras transações, a decisão do STJ abre a oportunidade de reaver valores pagos indevidamente nos últimos cinco anos, desde que respeitado o prazo prescricional.

O momento exige atenção redobrada: a busca de orientação jurídica especializada é fundamental para identificar eventuais distorções e acionar os meios adequados para resguardar direitos frente ao Fisco Municipal.