Publicada no DJE n.216, de 19/11/2025 p.93/94

NOTA TÉCNICA N. 02/2025-CGIA/TJRO

Segurança de sistemas de Inteligência Artificial no contexto judiciário: prevenção e controle de manipulação maliciosa de comandos (*prompt injection*) em documentos.

1. CONTEXTUALIZAÇÃO

O Tribunal de Justiça do Estado de Rondônia (TJRO), em conformidade com a Resolução n. 615/2025-CNJ, de 11 de março de 2025, a Resolução n. 335/2020-CNJ, de 29 de setembro de 2020, e a Política de Inteligência Artificial do TJRO (Resolução n. 356/2025-TJRO, de 28 de agosto de 2025), apresenta análise técnica sobre a vulnerabilidade conhecida como *prompt injection*, reconhecida na literatura especializada como risco relevante para sistemas judiciais automatizados (Badaró & Puppe, 2025; IBM, 2025).

2. DEFINIÇÃO TÉCNICA

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.

3. FUNDAMENTAÇÃO JURÍDICA E ÉTICA

3.1. Potenciais Repercussões no Ordenamento Jurídico

A prática pode caracterizar violação a:

a) princípios da boa-fé processual e dever de cooperação, previstos no Código de Processo Civil (Lei n. 13.105/2015), art. 5º, e disposições sobre litigância de má-fé, arts. 77, I a VI; 80; 81;

b) deveres profissionais dispostos no Estatuto da Advocacia (Lei n. 8.906/1994), art. 32, VIII, e art. 34, XIII;

c) deveres éticos previstos no Código de Ética e Disciplina da OAB (Resolução n. 02/2015-OAB, de 19 de outubro de 2015), art. 2º, parágrafo único, VI, e art. 32;

d) direito à revisão de decisões automatizadas e princípios da transparência e não discriminação, conforme a Lei Geral de Proteção de Dados Pessoais (Lei n. 13.709/2018), art. 20, § 1º, e art. 6º

e) Resolução CNJ nº 615/2025, art. 3º, que estabelece parâmetros de segurança e auditoria algorítmica.

3.2. Implicações Constitucionais

A manipulação de sistemas de IA judicial pode comprometer:

I - o devido processo legal (Constituição Federal, art. 5º, LIV);

II - o contraditório e a ampla defesa (Constituição Federal, art. 5º, LV);

III - a isonomia processual (Constituição Federal, art. 5º, caput);

IV - a duração razoável do processo (Constituição Federal, art. 5º, LXXVIII).

4. RISCOS IDENTIFICADOS

4.1 Risco à Imparcialidade

Sistemas de IA podem gerar resumos enviesados, classificações incorretas ou análises distorcidas, produzindo viés informacional, mesmo que a decisão do magistrado permaneça autônoma;

4.2 Risco à Segurança da Informação

Instruções maliciosas podem comprometer bases de dados jurisprudenciais e sistemas de gestão processual;

4.3 Risco à Confiabilidade Institucional

A exploração dessas vulnerabilidades pode abalar a confiança pública nos sistemas digitais do Judiciário.

5. MEDIDAS DE CONTROLE E MITIGAÇÃO

5.1 Técnicas

São requisitos técnicos para os sistemas de IA desenvolvidos no âmbito do Tribunal de Justiça do Estado de Rondônia, ou que venham a ser contratados:

I – Registro de logs detalhados: Assegurar que os sistemas de IA mantenham registro de todas as interações, incluindo data, hora, identificação do usuário, conteúdo dos *prompts* submetidos e respostas geradas, conforme estabelecido nos arts. 3º, XIII, e 10 dos Atos n. 1629/2025 e n. 1630/2025.

II – Validação de formato documental: Estabelecer processo de conversão de documentos para texto plano antes do processamento por IA, reduzindo a superfície de ataque através da remoção automática de metadados.

III – Testes durante homologação: Incluir nos procedimentos previstos no art. 14 da Resolução n. 356/2025-TJRO casos de teste específicos com documentos contendo técnicas conhecidas de *prompt injection*, verificando a resiliência dos sistemas antes de sua entrada em produção.

IV – Revisão periódica de outputs: Implementar amostragem estatística dos resultados gerados pela IA para identificação de inconsistências ou padrões atípicos, sem necessidade de análise exaustiva de todos os documentos processados.

6. Procedimentais

I – Revisão humana obrigatória de todos os resultados gerados por IA, conforme arts. 3º e 9º dos Atos n. 1629/2025 e n. 1630/2025;

II – proibição do uso de classificações automatizadas sem possibilidade de revisão humana;

III – Capacitação contínua dos servidores e magistrados sobre os riscos e protocolos relacionados.

7. Responsabilização

O TJRO poderá adotar as seguintes medidas administrativas e processuais em casos de identificação de manipulação de comandos em documentos processuais:

I – Apuração técnica e perícia documental em caso de suspeita de manipulação;

II – Possibilidade de adoção de sanções processuais cabíveis e comunicação aos órgãos competentes;

III – Encaminhamento de indícios de ilícitos aos órgãos responsáveis pela apuração, conforme a natureza da conduta.

8. RECOMENDAÇÕES

8.1. Divulgação e Transparência

I – Divulgar amplamente a presente Nota Técnica junto à comunidade jurídica e à sociedade.

II – Comunicar à Ordem dos Advogados do Brasil – Seccional Rondônia sobre os riscos identificados e as medidas de controle implementadas.

8.2 Orientação a Magistrados e Servidores

I – Alertar magistrados e servidores para observarem possíveis discrepâncias entre o conteúdo dos autos e os resultados de análises geradas por sistemas de IA, em conformidade com o dever de revisão estabelecido nos arts. 3º e 9º dos Atos n. 1629/2025 e n. 1630/2025.

II – Incluir o tema *prompt injection* nas capacitações continuadas previstas no art. 11 da Resolução n. 356/2025-TJRO.

III – Adotar as medidas processuais cabíveis quando identificada má-fé relacionada à manipulação de sistemas de IA, nos termos dos arts. 77, 80 e 81 do Código de Processo Civil, assegurado o contraditório e a ampla defesa.

IV – Determinar, quando couber, perícia técnica para identificação de técnicas de *prompt injection* em documentos processuais, conforme previsto no art. 156 do Código de Processo Civil.

8.3 Orientações à Secretaria de Tecnologia (STIC)

I – Incorporar nos procedimentos de homologação previstos no art. 14 da Resolução n. 356/2025-TJRO casos específicos de avaliação de vulnerabilidade a técnicas de *prompt injection*.

II – Aprimorar os mecanismos de monitoramento, auditoria e registro de logs das soluções de IA, conforme estabelecido no art. 18 da Resolução n. 356/2025-TJRO.

9. CONCLUSÃO

A integridade dos sistemas de inteligência artificial do Poder Judiciário é essencial para segurança, isonomia e confiança social. O *prompt injection* representa ameaça real à imparcialidade e à fiabilidade processual, exigindo respostas técnicas, normativas e éticas articuladas. A prática pode caracterizar violação a diversos deveres processuais e éticos, estando sujeita às consequências legais previstas no ordenamento jurídico.

10. REFERÊNCIAS

BADARÓ, Rodrigo; PUPPE, Matheus. Prompt injection jurisdicional: riscos e controles. OAB Nacional, 2025. Disponível em: https://www.oab.org.br/noticia/63465/artigo. Acesso em: 6 nov. 2025.

IBM Corporation. Prompt Injection Attacks: Threats and Safeguards. IBM Think, 2025. Disponível em: https://www.ibm.com/br-pt/think/topics/prompt-injection. Acesso em: 6 nov. 2025.

Desembargador Alexandre Miguel

Presidente do Comitê de Governança em Inteligência Artificial

Documento assinado eletronicamente por **ALEXANDRE MIGUEL**, **Presidente do Comitê**, em 17/11/2025, às 17:29 (horário de Rondônia), conforme § 3º do art. 4º, do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade do documento pode ser conferida no Portal SEI https://www.tjro.jus.br/sistema-eletronico-de-informacoes-sei, informando o código verificador **5246149** e o código CRC **73639CF0**.