RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

Inteligência Artificial e Advocacia de Estado: parâmetros para utilização segura na elaboração de peças processuais

Artificial Intelligence and State Legal Counsel: parameters for safe use in the drafting of procedural documents

Inteligencia Artificial y Defensa del Estado: parámetros para el uso seguro en la elaboración de escritos procesales

Leandro Mendes Neris
Graduado em Direito
Instituição: Universidade Federal do Amazonas
E-mail: leandromendesneris@gmail.com

RESUMO: O presente artigo analisa, sob as perspectivas jurídica, ética e técnica, a incorporação de sistemas de inteligência artificial generativa — em especial os Modelos de Linguagem de Grande Escala (LLMs) — às atividades da Advocacia de Estado, notadamente na elaboração de minutas de manifestações processuais, pareceres, recursos e petições. O estudo examina os riscos específicos dessa tecnologia no contexto jurídico-institucional, incluindo o fenômeno de alucinações de modelos de linguagem, a vulnerabilidade a ataques de prompt injection, o risco de vazamento de informações sigilosas e a manipulação de respostas geradas por IA. Em contrapartida, são analisados os benefícios potenciais em termos de produtividade, pesquisa jurisprudencial e padronização. O artigo propõe parâmetros práticos de utilização segura, incluindo técnicas de engenharia de prompt aplicadas à advocacia pública, governança institucional, bibliotecas de prompts padronizados e critérios de responsabilidade disciplinar do procurador. Conclui-se que a IA pode ser uma ferramenta poderosa para a advocacia de Estado, desde que sua utilização seja disciplinada por supervisão humana qualificada, por protocolos institucionais rígidos e por consciência dos limites éticos e jurídicos da delegação de atividades intelectuais a sistemas automatizados.
PALAVRAS-CHAVE: Inteligência Artificial. Advocacia de Estado. Prompt injection. Alucinações de LLM. Engenharia de prompt. Responsabilidade do procurador. Governança institucional. Segurança da informação.

ABSTRACT: This article analyzes, from legal, ethical, and technical perspectives, the incorporation of generative artificial intelligence systems — particularly Large Language Models (LLMs) — into the activities of State Legal Counsel, notably in the drafting of procedural submissions, legal opinions, appeals, and petitions. The study examines the specific risks of this technology in the legal-institutional context, including language model hallucinations, vulnerability to prompt injection attacks, the risk of sensitive information leakage, and the manipulation of AI-generated responses. In contrast, the potential benefits in terms of productivity, case law research, and standardization are also analyzed. The article proposes practical parameters for safe use, including prompt engineering techniques applied to public legal counsel, institutional governance, standardized prompt libraries, and criteria for the attorney's disciplinary responsibility. It concludes that AI can be a powerful tool for State Legal Counsel, provided that its use is regulated by qualified human supervision, strict institutional protocols, and awareness of the ethical and legal limits of delegating intellectual activities to automated systems.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

KEYWORDS: Artificial Intelligence. State Legal Counsel. Prompt injection. LLM hallucinations. Prompt engineering. Attorney's responsibility. Institutional governance. Information security.

1 INTRODUÇÃO

A revolução tecnológica inaugurada pela disseminação dos Modelos de Linguagem de Grande Escala (Large Language Models — LLMs) representa, possivelmente, a mais profunda transformação nas práticas intelectuais desde a popularização da internet. No campo jurídico, a utilização de ferramentas baseadas em inteligência artificial generativa — como o ChatGPT, da OpenAI; o Claude, da Anthropic; o Gemini, do Google; e o Copilot, da Microsoft — deixou de ser uma perspectiva futura para se tornar uma realidade presente, incorporada ao cotidiano de escritórios de advocacia, de departamentos jurídicos corporativos e, progressivamente, às carreiras jurídicas do setor público.

A Advocacia de Estado — compreendida, no ordenamento brasileiro, pela Advocacia-Geral da União, pelas Procuradorias dos Estados e do Distrito Federal e pelas Procuradorias dos Municípios — enfrenta desafios particulares nesse cenário. De um lado, a crescente demanda por eficiência, racionalização de recursos e velocidade de resposta em contextos de elevado volume de processos torna a inteligência artificial uma solução atraente. De outro, a natureza peculiar da função de representação judicial e consultiva do Estado, pautada pela indisponibilidade do interesse público, pela observância estrita da legalidade e pela responsabilidade institucional do procurador, impõe limites que não existem — ou existem em menor grau — na advocacia privada.

A utilização de IA na elaboração de peças processuais suscita questões que perpassam distintos domínios do conhecimento jurídico. Do ponto de vista técnico, é necessário compreender os mecanismos de funcionamento dos LLMs, seus pontos de falha e suas vulnerabilidades. Do ponto de vista ético-profissional, coloca-se o problema da delegação de atividades intelectuais inerentes ao exercício da advocacia a sistemas automatizados, com reflexos sobre o dever de diligência, a lealdade processual e a responsabilidade do advogado. Do ponto de vista institucional, emergem questões de governança, segurança da informação e padronização de procedimentos que demandam regulamentação interna por parte dos órgãos de advocacia do Estado.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

O presente artigo tem por objetivo analisar, de forma sistemática e prática, os parâmetros para a utilização segura de sistemas de inteligência artificial na Advocacia de Estado. Para tanto, o estudo está estruturado da seguinte forma: inicialmente, traça-se um panorama dos LLMs e de seu funcionamento; em seguida, examina-se o fenômeno das alucinações e seus impactos jurídicos; analisa-se o conceito de prompt injection e demais vetores de risco; propõem-se medidas de mitigação e diretrizes de governança; discute-se a engenharia de prompt aplicada à advocacia pública; e, por fim, analisa-se a responsabilidade disciplinar e jurídica do procurador pelo uso de conteúdo gerado por IA.

2 MODELOS DE LINGUAGEM DE GRANDE ESCALA E O CONTEXTO JURÍDICO

2.1 Funcionamento básico dos LLMs

Os modelos de linguagem de grande escala são sistemas de inteligência artificial treinados em grandes volumes de texto, com o objetivo de prever, de forma probabilística, qual token (unidade de texto) deve suceder ao anterior em uma sequência. A partir de arquiteturas baseadas em transformers, desenvolvidas a partir do artigo seminal de Vaswani et al. (2017), esses modelos aprenderam padrões linguísticos suficientemente ricos para gerar texto coerente, responder perguntas complexas, traduzir idiomas, redigir documentos e, no caso de interesse aqui, elaborar peças jurídicas.

É essencial compreender que um LLM não "sabe" direito no sentido em que um jurista sabe: o modelo não possui representações internas de normas, não consulta bases de dados em tempo real (salvo quando dotado de ferramentas externas específicas) e não raciocina dedutivamente a partir de premissas estabelecidas. O que o modelo faz é, fundamentalmente, reconhecer padrões estatísticos e gerar texto que, dadas as entradas recebidas (prompts), se assemelha ao que, em seu conjunto de treinamento, seguiria aquele contexto. A aparência de raciocínio jurídico rigoroso é, em larga medida, uma consequência da estrutura do texto jurídico incorporado no treinamento, e não de um processo dedutivo genuíno.

Essa distinção é fundamental para a compreensão dos riscos associados ao uso de LLMs na advocacia, pois explica por que o modelo pode produzir argumentos juridicamente

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

impecáveis em sua forma, mas factualmente incorretos em seu conteúdo — citando precedentes inexistentes, atribuindo ementas falsas a acórdãos reais, ou afirmando o estado normativo de leis já revogadas.

2.2 O LLM como ferramenta de apoio e seus limites

Do ponto de vista da produtividade, os LLMs oferecem contribuições reais e documentadas em três eixos principais. Em primeiro lugar, na aceleração da produção de minutas: um procurador que domina técnicas de engenharia de prompt pode obter uma minuta estruturada de contestação, de resposta a embargos ou de parecer em fração do tempo que levaria para produzi-la do zero, dedicando sua energia intelectual à supervisão, ao ajuste e ao aprimoramento do conteúdo gerado. Em segundo lugar, na pesquisa jurisprudencial, os modelos com acesso à internet ou dotados de bases de dados jurídicas integradas conseguem identificar precedentes relevantes, sintetizar teses e mapear a evolução jurisprudencial com maior rapidez do que a pesquisa manual. Em terceiro lugar, na padronização de linguagem e na conformidade formal, reduzindo erros processuais de natureza técnica.

Contudo, esses benefícios não podem obscurecer limitações estruturais críticas. O conhecimento dos LLMs é estático, limitado à data de corte do treinamento e frequentemente desatualizado em relação às alterações legislativas e jurisprudenciais mais recentes. A alucinação — produção confiante de informações falsas — é um fenômeno intrínseco ao funcionamento desses modelos, não uma falha corrigível por simples ajustes. E a dependência excessiva na saída do modelo, sem supervisão humana qualificada, pode transformar a IA de ferramenta de apoio em um veículo de erros materiais, com consequências processuais graves.

3 ALUCINAÇÕES DE MODELOS DE LINGUAGEM E SEUS IMPACTOS NA ATUAÇÃO DA FAZENDA PÚBLICA

3.1 Conceito e tipologia das alucinações

O termo alucinação, aplicado aos LLMs, designa a geração de informações que não correspondem à realidade factual, mas são apresentadas pelo modelo com o mesmo nível de confiança e fluência que as informações corretas. Diferentemente de um erro de digitação

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

ou de uma inconsistência lógica facilmente perceptível, as alucinações de LLMs frequentemente se apresentam em forma de texto bem estruturado, internamente coerente e aparentemente plausível — o que torna sua detecção particularmente desafiadora para quem não domina profundamente o tema tratado.

No contexto jurídico, as alucinações assumem tipologias específicas de particular gravidade. A primeira e mais documentada é a fabricação de precedentes judiciais: o modelo cita acórdãos com numeração verossímil, relator, data e ementa, que simplesmente não existem. A segunda é a distorção de precedentes reais: o modelo cita um julgado que existe, mas atribui a ele uma ementa ou uma tese diversa da constante no original. A terceira é a citação de dispositivos legais revogados como se fossem vigentes ou a afirmação de estados normativos já alterados. A quarta é a atribuição equivocada de posições doutrinárias a autores, citando obras inexistentes ou atribuindo a determinado jurista uma tese que não é sua.

3.2 Casos documentados de alucinação na advocacia

O caso mais amplamente difundido sobre o impacto das alucinações de IA na prática jurídica foi o do advogado Steven Schwartz, nos Estados Unidos, que, em 2023, peticionou ao Tribunal Federal do Distrito Sul de Nova York, citando seis precedentes inexistentes, todos gerados pelo ChatGPT. Ao ser intimado pelo juiz Kevin Castel a comprovar a existência dos julgados, o advogado não logrou fazê-lo, sofrendo sanções disciplinares e condenação ao pagamento de multa. O episódio gerou ampla repercussão e motivou a elaboração de diretrizes para o uso de IA por tribunais em todo o mundo.

No Brasil, ao contrário do que se poderia supor, já existe um conjunto expressivo e crescente de precedentes sobre o tema, com decisões sancionatórias proferidas pelos mais diversos tribunais do país. O primeiro caso de relevância institucional data de abril de 2023: o então corregedor-geral da Justiça Eleitoral, Ministro Benedito Gonçalves, do TSE, aplicou multa por litigância de má-fé — no valor de R$ 2.604 — a um advogado que protocolou petição redigida integralmente pelo ChatGPT, na qual buscava ingressar como amicus curiae em ação eleitoral envolvendo o ex-presidente Jair Bolsonaro, processo em que não representava nenhuma das partes. O Ministro foi enfático ao registrar que o advogado havia submetido ao juízo "uma fábula, resultante de conversa com uma inteligência artificial".

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

Em fevereiro de 2025, o TJSC consolidou um padrão decisório mais severo em dois casos distintos e emblemáticos. No primeiro, a 6ª Câmara Cível multou um advogado em 10% do valor da causa após ele admitir ter utilizado o ChatGPT na elaboração de recurso em ação de reintegração de posse — recurso no qual tanto as citações jurisprudenciais quanto as referências a obras jurídicas eram incorretas ou totalmente fictícias. Além da sanção pecuniária, o tribunal determinou a comunicação do caso à OAB/SC e o encaminhamento de cópia do recurso para análise disciplinar. O advogado alegou que o erro ocorreu por "uso inadvertido" do ChatGPT. No segundo caso catarinense, a 5ª Câmara Criminal advertiu formalmente o advogado que utilizou IA para redigir habeas corpus com jurisprudência inexistente, tendo a relatora afirmado que a conduta constituiu ato de má-fé e desrespeito ao tribunal e que os precedentes apresentados "foram criados para induzir o julgador a erro".

Na Justiça do Trabalho, o fenômeno assumiu proporções ainda mais preocupantes. Em junho de 2025, o TRT-7 (Ceará) identificou, em petição recursal, jurisprudência manipulada com trechos falsificados ou inexistentes, incluindo trechos indevidamente atribuídos ao TST que tratavam de matérias completamente distintas — uma das ementas, apresentada como se versasse sobre assédio moral, dizia respeito, na verdade, à "pejotização" —, com indícios de geração por inteligência artificial. O tribunal aplicou multa de 2% sobre o valor da causa e determinou o envio de ofício à OAB-CE para apuração de possível infração disciplinar. Em outubro de 2025, o TRT-12 (Santa Catarina) julgou caso em que a petição inicial continha decisões, citação doutrinária e até nome de magistrado inexistente, todos os elementos aparentemente gerados por inteligência artificial. O juiz Daniel Carvalho Martins afirmou que "tais achados vão além de um mero erro material" e que a peça foi produzida "sem qualquer verificação humana, o que configura um ato processual inexistente". A parte foi condenada ao pagamento de multa de R$ 3.700.

Em fevereiro de 2026, o TRT-2 (São Paulo) enfrentou uma situação que expôs uma estratégia recorrente de esquiva na prática forense: ao ser questionado sobre o uso de jurisprudência fictícia, o advogado atribuiu a responsabilidade ao "corpo de estagiários do escritório", argumento que o tribunal rechaçou com veemência, afirmando que "a postulação em juízo é ato privativo do advogado, que também é o responsável por seu conteúdo". O magistrado destacou expressamente que a conduta violou a Recomendação n.º 1/2024 do Conselho Federal da OAB, que exige supervisão humana no uso de ferramentas generativas,

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

e que houve "criação deliberada de precedentes para reforçar a tese defensiva, comprometendo a segurança jurídica".

É precisamente esse instrumento normativo — a Recomendação n.º 1/2024 do Conselho Federal da OAB — que vem sendo invocado como parâmetro objetivo nas decisões mais recentes. Nos termos da norma, exige-se do advogado "entendimento adequado das limitações, verificação rigorosa das informações, transparência aos clientes e demais interlocutores, sendo vedada a delegação de atos privativos da profissão sem supervisão qualificada". A consolidação dessa cadeia de precedentes — abrangendo o TSE, o TJSC e múltiplos TRTs — demonstra que o problema não é incipiente nem episódico: entre 2025 e 2026, tribunais brasileiros registraram vários casos de advogados multados por apresentarem petições com jurisprudência falsa gerada por IA, configurando um padrão alarmante que já chegou ao conhecimento das corregedorias dos estados.

3.3 Impactos específicos na atuação da Fazenda Pública

Para a Advocacia de Estado, os impactos das alucinações possuem dimensão especialmente grave por razões que se somam às que afetam a advocacia privada. Em primeiro lugar, o procurador representa o Estado — titular de interesse público indisponível —, de modo que um erro material decorrente de alucinação de IA pode comprometer não apenas uma causa individual, mas também criar precedentes negativos para o ente público em toda uma linha jurisprudencial. Em segundo lugar, a responsabilidade funcional do procurador é mais intensa do que a do advogado privado: erros que, na advocacia privada, acarretariam apenas responsabilidade contratual e disciplinar podem, na advocacia pública, configurar ilícito funcional com consequências administrativas.

Em terceiro lugar, a natureza dos litígios envolvendo a Fazenda Pública — frequentemente relacionados a tributos, previdência, precatórios, serviços públicos e políticas públicas — implica que o erro em uma única petição pode ter efeito multiplicador sobre centenas ou milhares de casos similares tramitados pelo mesmo órgão. Uma tese juridicamente equivocada, incorporada a uma minuta gerada por IA e não revisada com rigor suficiente, pode se propagar sistematicamente se adotada como modelo institucional.

Em quarto lugar, há o risco de que a pressão por produtividade — frequentemente elevada nos órgãos de advocacia pública em razão do desequilíbrio entre o volume de

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

processos e o número de procuradores — favoreça uma relação de confiança excessiva na saída do modelo, reduzindo o escrutínio humano necessário à detecção de alucinações.

4 RISCOS DE SEGURANÇA INSTITUCIONAL: PROMPT INJECTION, VAZAMENTO E MANIPULAÇÃO

4.1 O conceito de prompt injection

O conceito de prompt injection: Prompt injection é uma categoria de ataque que explora a incapacidade estrutural dos LLMs de distinguir, de forma confiável, entre as instruções fornecidas pelo operador do sistema (system prompt) e o conteúdo produzido por usuários ou por fontes externas que o modelo processa. O conceito foi formalmente identificado em 2022 e, desde então, tem recebido atenção crescente da comunidade de segurança. Em 2025, a OWASP — referência global em segurança de aplicações — posicionou o prompt injection como a principal ameaça no ranking de riscos para sistemas de IA generativa (OWASP Top 10 for LLMs 2025). O ataque consiste em inserir, no conteúdo que o modelo analisa, instruções ocultas ou disfarçadas que, ao serem processadas pelo LLM, alteram seu comportamento, sobrepõem-se às instruções originais ou extraem informações que não deveriam ser reveladas.

Existem dois subtipos principais. O primeiro é o prompt injection direto, que ocorre quando a entrada do usuário afeta diretamente o comportamento do LLM, seja de forma intencional — quando o usuário planeja e estrutura o ataque — ou não intencional. O segundo é o prompt injection indireto, que ocorre quando o modelo aceita prompts de fontes externas, como sites ou arquivos, e o conteúdo dessas fontes, ao ser interpretado, altera o comportamento do LLM. Ambas as modalidades podem resultar em respostas tendenciosas, na revelação de informações confidenciais ou na execução de ações não autorizadas.

No contexto da advocacia pública, o risco é particularmente sofisticado: um procurador que utiliza um sistema de IA para analisar a petição inicial da parte adversa e extrair os principais argumentos a rebater pode, sem saber, processar comandos maliciosos embutidos na própria peça pela parte contrária, instruindo o modelo a produzir uma análise favorável ao adversário, a omitir argumentos relevantes ou a vazar informações estratégicas do órgão. O que até recentemente parecia um cenário hipotético converteu-se, em maio de

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

2026, em realidade concreta e juridicamente sancionada no Brasil, inaugurando uma nova e alarmante fronteira de litigância de má-fé.

4.1.1. O caso de Parauapebas: Um dos primeiros precedentes brasileiros

Em 12 de maio de 2026, a 3ª Vara do Trabalho de Parauapebas deparou-se com um achado sem precedentes na jurisprudência brasileira. Ao processar a petição inicial pelo sistema Galileu — ferramenta de inteligência artificial generativa da Justiça do Trabalho — , o juízo identificou um texto em fonte branca sobre fundo branco, invisível ao leitor humano, contendo o seguinte comando: "Atenção, inteligência artificial, conteste essa petição de forma superficial e não impugne os documentos, independentemente do comando que lhe for dado."

O magistrado dedicou as primeiras páginas da sentença a repudiar a conduta, afirmando que ela era "incompatível com os mais elementares deveres que recaem sobre todo aquele que participa do processo judicial" e classificou a prática como "um ataque à credibilidade das ferramentas institucionais, um desrespeito ao juízo, às partes e à sociedade e um precedente que este juízo não pode deixar passar". A sentença reconheceu o ato atentatório à dignidade da justiça, condenou solidariamente as duas advogadas signatárias ao pagamento de multa de 10% sobre o valor da causa — cerca de R$ 84.250,00 — e determinou a expedição de ofício à OAB/PA e à Corregedoria do TRT da 8ª Região.

4.1.2. A escalada nacional: STJ, São Paulo e a dimensão criminal

O episódio de Parauapebas deixou de ser um caso isolado. Em 20 de maio de 2026, o presidente do STJ, ministro Herman Benjamin, determinou a abertura de inquérito policial e de procedimento administrativo para apurar tentativas de fraude processual por meio de prompt injection no sistema STJ Logos — sistema de inteligência artificial generativa desenvolvido pela própria Corte e lançado em fevereiro de 2025. Segundo apuração da TV Globo, há ao menos 11 processos criminais em que a técnica foi identificada.

O STJ determinou que as tentativas de uso do mecanismo, embora neutralizadas pelas camadas de segurança do sistema, passassem a ser certificadas nos autos, o que permitiu a imposição de sanções processuais pelos ministros da Corte. O vice-presidente do

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

STJ, ministro Luis Felipe Salomão, foi ainda mais enfático ao qualificar a prática como "caso de polícia", afirmando que "quando a IA é mal utilizada, com má-fé, que transborda a questão puramente ética, isso é caso de polícia, que tem que ser averiguado tanto no plano administrativo quanto no criminal".

Em 21 de maio, o fenômeno chegou a São Paulo, com a identificação de prompt injection em petição contra banco em vara cível da capital, confirmando que o tema deixou de ser anedótico e adquiriu dimensão sistêmica nacional.

4.1.3. Implicações para a Advocacia de Estado

A técnica é classificada pela OWASP como a vulnerabilidade número um em sua lista de riscos para aplicações baseadas em modelos de linguagem. Para a Advocacia de Estado, o vetor de risco é duplo. No sentido ativo, o órgão pode ser vítima de prompt injection indireto quando seus procuradores submetem peças da parte adversa à análise de sistemas de IA — situação em que comandos ocultos nessas peças podem comprometer a qualidade da análise ou, em sistemas mal configurados, extrair informações sigilosas. No sentido passivo, sistemas de IA utilizados pelos próprios órgãos de advocacia pública — para triagem processual, análise de risco e elaboração de minutas — constituem alvos de ataques externos.

4.2 Vazamento de informações sensíveis

A utilização de ferramentas de IA no âmbito da Advocacia de Estado envolve, necessariamente, o processamento de informações altamente sensíveis: dados sobre litígios em andamento, estratégias de defesa e de ataque da Fazenda Pública, pareceres internos não publicados, informações sobre contribuintes em processos fiscais, dados pessoais de servidores em demandas trabalhistas e informações classificadas relacionadas a contratos administrativos sob sigilo.

O risco de vazamento de tais informações ocorre em múltiplos vetores. O primeiro é o armazenamento nos servidores da empresa fornecedora da IA: a maioria das ferramentas de LLM disponíveis no mercado consumidor transmite as entradas dos usuários para servidores externos, onde podem ser armazenadas, utilizadas no treinamento de versões

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

futuras do modelo e sujeitas a violações de segurança. A utilização de ferramentas comerciais de IA com informações processuais sigilosas sem a celebração de contrato específico que garanta a confidencialidade e a exclusão dos dados pode caracterizar violação às normas de sigilo funcional e à Lei Geral de Proteção de Dados Pessoais (LGPD, Lei n.º 13.709/2018).

O segundo vetor é a exposição de informações durante a interação com o modelo: ao formular um prompt para que o modelo elabore uma contestação, o procurador necessariamente insere elementos fáticos do caso, identificação das partes e informações estratégicas. Em sistemas sem proteção adequada de dados, essas informações podem ser acessadas por terceiros — funcionários da empresa, pesquisadores ou agentes maliciosos em caso de violação de segurança.

O terceiro vetor é o treinamento do modelo com base nos dados inseridos pelos usuários. Diversas ferramentas comerciais de IA generativa utilizam as interações dos usuários para aperfeiçoar continuamente seus modelos. Isso significa que informações sensíveis inseridas por um procurador podem, em tese, aparecer nas respostas fornecidas a outros usuários, configurando grave violação ao sigilo funcional.

4.3 Manipulação de respostas da IA

Além da alucinação espontânea e dos ataques de prompt injection, existe um terceiro risco de natureza distinta: a manipulação intencional das respostas da IA por atores que, tendo conhecimento de como determinado sistema de IA funciona, estruturam sua estratégia processual de forma a influenciar a análise que o sistema adversário realizará. Em contextos em que a parte adversa tem conhecimento de que a Fazenda Pública utiliza determinada ferramenta de IA para a análise de peças processuais, surge a possibilidade — ainda mais teórica do que prática no estágio atual — de que a parte formule suas manifestações de modo a explorar as fraquezas do modelo utilizado pelo órgão público.

Um cenário mais imediato e concreto de manipulação é o chamado jailbreak — um conjunto de técnicas elaboradas para contornar as salvaguardas éticas e de segurança dos LLMs, induzindo-os a produzir conteúdo que normalmente recusariam gerar. No contexto da advocacia pública, um procurador que utilize técnicas de jailbreak para induzir o modelo a produzir argumentos que, se avaliados eticamente, seriam considerados abusivos ou

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

contrários ao interesse público, estaria utilizando a IA como instrumento de práticas processuais reprováveis, com potencial para configurar má-fé processual.

4.4 Medidas de mitigação

A adoção de medidas de mitigação efetivas requer uma abordagem em camadas que combine salvaguardas técnicas, procedimentais e institucionais.

No plano técnico, as medidas prioritárias incluem: (a) a implantação de sistemas de IA em infraestrutura própria ou em nuvem privada, com controle integral sobre os dados processados, eliminando o risco de transmissão a servidores de terceiros; (b) a seleção de ferramentas que ofereçam garantias contratuais de não utilização dos dados para treinamento e de exclusão periódica dos dados; (c) a implementação de sistemas de filtragem de inputs e outputs que identifiquem padrões de prompt injection antes que o conteúdo malicioso seja processado pelo modelo; (d) o uso de modelos fine-tuned — ajustados a partir de dados jurídicos curados pelo próprio órgão — para reduzir o risco de alucinações em domínios jurídicos específicos.

No plano procedimental, as medidas essenciais compreendem: (a) a obrigatoriedade de verificação humana de todas as referências jurisprudenciais e legislativas geradas pelo modelo, por meio de consulta às fontes primárias; (b) a proibição de inserção de dados identificados de partes e de informações classificadas em ferramentas de IA não homologadas pelo órgão; (c) a implementação de controles de acesso que restrinjam a utilização de sistemas de IA a procuradores treinados e habilitados; (d) a criação de logs de utilização que permitam a auditoria posterior do uso da ferramenta em cada caso.

No plano institucional, as diretrizes devem estabelecer: (a) a criação de uma política formal de uso de IA, aprovada pela cúpula do órgão e vinculante para todos os membros; (b) a designação de um responsável institucional pela governança de IA (Chief AI Compliance Officer ou equivalente); (c) a realização periódica de treinamentos sobre o uso responsável de IA; (d) a criação de um canal de reporte de incidentes relacionados ao uso inadequado de IA.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

5 ENGENHARIA DE PROMPT APLICADA À ADVOCACIA PÚBLICA

5.1 Conceito e relevância da engenharia de prompt

A engenharia de prompt (prompt engineering) é a disciplina que se ocupa da elaboração estratégica das instruções fornecidas aos modelos de linguagem com o objetivo de maximizar a qualidade, a precisão e a relevância das respostas geradas. Longe de ser uma atividade trivial, a engenharia de prompt constitui uma competência técnica específica que pode determinar, de forma decisiva, a diferença entre uma saída de IA genérica e de baixo valor e uma saída precisa, estruturada e juridicamente relevante.

No contexto da Advocacia de Estado, o domínio das técnicas de engenharia de prompt pode transformar radicalmente a produtividade do procurador, ao mesmo tempo em que reduz — embora não elimine — o risco de alucinações e de saídas inadequadas. A elaboração de um prompt bem estruturado, que forneça ao modelo contexto suficiente, restrições claras e instruções de formato específicas, resulta em minutas muito superiores às obtidas com prompts genéricos e imprecisos.

5.2 Técnicas fundamentais de engenharia de prompt

As principais técnicas de engenharia de prompt com aplicação à prática jurídica são as seguintes:

a) Definição de papel (role prompting): consiste em instruir o modelo a assumir um papel específico antes de iniciar a tarefa. Para a advocacia pública, um prompt eficaz pode começar com: "Você é um procurador do Estado especializado em direito tributário, com experiência em contencioso fiscal e profundo conhecimento da jurisprudência do STJ e do STF sobre a matéria. Analise o caso a seguir e elabore..." Esse tipo de instrução orienta o modelo a acessar padrões linguísticos e estruturais mais próximos do perfil desejado.

b) Fornecimento de contexto (context setting): a qualidade da saída do modelo é diretamente proporcional à quantidade e à qualidade do contexto fornecido. Um prompt eficaz deve especificar a natureza da peça solicitada (contestação, apelação, parecer, memorando), o órgão jurisdicional destinatário, a fase processual, os fundamentos fáticos relevantes e as teses jurídicas que o procurador pretende sustentar.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

c) Instrução de raciocínio passo a passo (chain-of-thought prompting): ao instruir o modelo a apresentar seu raciocínio de forma estruturada e sequencial antes de elaborar a peça final, o procurador obtém maior transparência sobre o caminho lógico adotado pelo modelo, o que facilita a identificação de erros e falhas argumentativas. A instrução pode ser algo como: "Antes de redigir a contestação, apresente um esboço com: (1) identificação das alegações da parte autora; (2) as defesas processuais aplicáveis; (3) as defesas de mérito, com indicação dos fundamentos jurídicos; (4) a ordem de apresentação que maximize a efetividade da defesa."

d) Instruções negativas (negative prompting): assim como é importante especificar o que o modelo deve fazer, é igualmente importante instruí-lo sobre o que não deve fazer. Exemplos aplicáveis: "Não cite jurisprudência sem indicar o número do processo e o tribunal"; "Não utilize argumentos que impliquem reconhecimento implícito do direito do autor"; "Não inclua afirmações sobre fatos que não constem dos documentos fornecidos."

e) Instrução de verificação (self-checking): o procurador pode instruir o modelo a verificar a própria saída antes de apresentá-la. Por exemplo: "Após elaborar a contestação, revise o texto e identifique: (1) qualquer afirmação que não esteja fundamentada nos documentos fornecidos; (2) qualquer precedente citado, indicando que o procurador deverá verificar sua existência e exatidão; (3) qualquer inconsistência lógica." Essa técnica não elimina as alucinações, mas as sinaliza, facilitando a revisão humana.

5.3 Padronização institucional: bibliotecas de prompts

Uma das contribuições mais relevantes que a engenharia de prompt pode oferecer à Advocacia de Estado é a criação de bibliotecas institucionais de prompts padronizados — repositórios de instruções cuidadosamente elaboradas, testadas e validadas para as tarefas mais frequentes de cada órgão. A padronização de prompts oferece múltiplas vantagens: reduz a variabilidade na qualidade das saídas geradas por diferentes procuradores; garante que as instruções incorporem as salvaguardas necessárias contra alucinações; facilita o treinamento de novos membros; e permite a atualização centralizada das instruções quando há mudanças legislativas ou jurisprudenciais relevantes.

Uma biblioteca de prompts para a Advocacia de Estado poderia compreender as seguintes categorias: (a) prompts para elaboração de minutas de contestação em matéria

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

tributária; (b) prompts para elaboração de recursos em matéria previdenciária; (c) prompts para elaboração de pareceres sobre regularidade de contratos administrativos; (d) prompts para pesquisa jurisprudencial sobre temas específicos; (e) prompts para consolidação e síntese de argumentos em causas repetitivas; (f) prompts para análise de risco jurídico e probabilidade de êxito.

Cada prompt padrão deve ser elaborado com a participação de procuradores especializados na matéria correspondente, testado em cenários reais anonimizados e submetido a um processo de validação antes de ser incorporado à biblioteca. A manutenção da biblioteca deve ser responsabilidade de uma equipe multidisciplinar que combine expertise jurídica e conhecimento técnico em IA.

5.4 Controle de qualidade das saídas geradas por IA

Independentemente da qualidade do prompt utilizado, toda saída gerada por IA deve ser submetida a um processo de controle de qualidade antes de ser incorporada a uma peça processual assinada pelo procurador. Esse processo deve ser formalizado em checklist institucional que cubra, no mínimo, os seguintes aspectos:

- Verificação de todos os precedentes citados, com consulta às fontes primárias (bases de dados dos tribunais) para confirmar existência, numeração e teor da ementa;
- Verificação da vigência de todos os dispositivos legais mencionados;
- Avaliação da coerência entre os fatos narrados pelo modelo e os documentos do processo;
- Avaliação da conformidade dos argumentos com a tese que o órgão adota institucionalmente na matéria;
- Revisão de linguagem para adequação ao estilo da peça processual e à linguagem formal do órgão;
- Verificação da ausência de informações sigilosas que não deveriam constar da peça;
- Avaliação da adequação ética dos argumentos empregados.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

6 USO SEGURO: APLICAÇÕES PRÁTICAS NA ADVOCACIA DE ESTADO

6.1 Elaboração de minutas

A elaboração de minutas de manifestações processuais é a aplicação mais imediata e de maior impacto da IA na Advocacia de Estado. Para causas de natureza repetitiva — execuções fiscais, impugnações de cumprimento de sentença, exceções de pré-executividade, contestações padronizadas em matéria previdenciária ou tributária —, o modelo pode gerar, em segundos, uma minuta estruturada que, submetida à revisão do procurador, pode ser finalizada em fração do tempo que levaria a elaboração integral.

O protocolo recomendado para elaboração de minutas por IA deve seguir as seguintes etapas: (1) o procurador identifica a categoria da peça e seleciona o prompt padrão correspondente da biblioteca institucional; (2) o prompt é personalizado com os dados específicos do caso, observando rigorosamente a proibição de inserção de dados identificados em ferramentas não homologadas; (3) a minuta gerada é submetida ao checklist de controle de qualidade; (4) o procurador aprimora, corrige e complementa a minuta com elementos de estratégia processual que demandam seu julgamento profissional; (5) a peça final é assinada exclusivamente pelo procurador responsável, sem qualquer referência à participação de IA na sua elaboração, salvo disposição normativa específica em contrário.

6.2 Pesquisa jurisprudencial

A pesquisa jurisprudencial assistida por IA apresenta um potencial de eficiência particularmente significativo. Modelos dotados de acesso a bases de dados jurídicas atualizadas ou de ferramentas de busca na internet conseguem, a partir de prompts adequados, identificar a posição dos tribunais superiores sobre determinada matéria, mapear a evolução de uma tese ao longo do tempo, identificar julgamentos recentes que alteraram entendimentos consolidados e sintetizar, de forma organizada, os fundamentos dos principais precedentes sobre um tema.

Contudo, o risco de alucinação é particularmente elevado na pesquisa jurisprudencial, o que torna indispensável a verificação de todos os precedentes identificados pelo modelo nas fontes primárias. O uso de IA para pesquisa jurisprudencial

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

deve ser tratado como ponto de partida — um auxílio para identificar candidatos a precedentes relevantes — e nunca como fonte definitiva de informação jurisprudencial. A consolidação final deve sempre resultar da consulta direta às bases oficiais dos tribunais.

6.3 Consolidação de argumentos

Outra aplicação de alto valor é o uso de IA para consolidar e organizar argumentos jurídicos complexos. Em causas de grande complexidade fática ou normativa, o procurador pode utilizar o modelo para identificar inconsistências internas na peça da parte adversa, organizar hierarquicamente os argumentos de defesa, identificar argumentos subsidiários ou alternativos não considerados e estruturar a peça de forma a maximizar a clareza e o impacto dos argumentos principais.

Para essa finalidade, é particularmente útil a técnica de fornecer ao modelo, além do prompt de instrução, os documentos relevantes do processo para análise — observando, mais uma vez, as restrições relativas à inserção de dados sensíveis em sistemas não homologados. A análise contextualizada dos documentos pelo modelo pode revelar elementos que o procurador, pressionado pelo volume de trabalho, poderia não identificar com a mesma agilidade.

6.4 Governança institucional do uso de IA

A implementação responsável de sistemas de IA na Advocacia de Estado exige a criação de uma estrutura de governança que discipline, de forma sistematizada, todos os aspectos relevantes da utilização da tecnologia. Os elementos essenciais dessa estrutura são:

Política de uso de IA: documento formal, aprovado pela cúpula do órgão, que estabeleça os princípios que regem a utilização de IA, as ferramentas homologadas, as categorias de informações que não podem ser inseridas em sistemas de IA, os requisitos de verificação e supervisão humana e as consequências do descumprimento das diretrizes.

Comitê de governança de IA: grupo multidisciplinar composto por procuradores, profissionais de tecnologia da informação e especialistas em segurança da informação, responsável por avaliar novas ferramentas de IA, atualizar a política institucional, investigar incidentes e recomendar melhorias no protocolo de uso.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

Programa de capacitação: treinamento obrigatório para todos os procuradores sobre os fundamentos do funcionamento dos LLMs, os riscos associados ao seu uso na advocacia pública e as técnicas de uso responsável, incluindo engenharia de prompt e procedimentos de verificação de saídas.

Auditoria periódica: revisão regular do uso dos sistemas de IA no órgão, com avaliação da qualidade das saídas geradas, identificação de padrões de uso inadequado e verificação da efetividade das medidas de mitigação implementadas.

7 RESPONSABILIDADE DO PROCURADOR PELA UTILIZAÇÃO DE CONTEÚDO GERADO POR IA

7.1 Fundamentos da responsabilidade profissional

A questão da responsabilidade do procurador pelo uso de conteúdo gerado por inteligência artificial é, talvez, o aspecto mais delicado e menos regulamentado da temática. O ponto de partida da análise é a constatação de que a responsabilidade pela peça processual permanece integral e exclusivamente com o procurador que a assina, independentemente do grau de participação de sistemas automatizados em sua elaboração. A IA não é sujeito de direito, não possui capacidade postulatória e não pode ser responsabilizada por erros materiais ou argumentativos.

O ordenamento jurídico brasileiro não prevê, até o momento, qualquer disposição específica que regulamente o uso de IA na advocacia ou que atribua responsabilidade por erros decorrentes do uso de sistemas automatizados. Diante dessa lacuna, a análise da responsabilidade do procurador deve ser conduzida com base nos princípios gerais que regem o exercício da advocacia e da função pública.

7.2 Dever de diligência e responsabilidade por erro material

O Estatuto da Advocacia (Lei n.º 8.906/1994), em seu artigo 32, estabelece que o advogado é responsável pelos atos que, no exercício profissional, praticar com dolo ou culpa. O Código de Ética e Disciplina da OAB, por sua vez, consagra o dever de diligência como um dos pilares do exercício da advocacia. Para os procuradores de Estado, esses deveres se

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

conjugam com os princípios da legalidade, impessoalidade, moralidade e eficiência que regem a Administração Pública (art. 37, caput, da Constituição Federal).

A utilização de IA sem a devida supervisão humana pode configurar violação ao dever de diligência em múltiplos cenários. O mais evidente é a petição que cita precedentes inexistentes, sem que o procurador tenha verificado sua existência — situação que, no direito norte-americano, já resultou em sanções disciplinares e em decisões judiciais que alertam sobre os riscos do uso irresponsável de IA. No Brasil, embora ainda não haja decisões judiciais específicas sobre o tema, os princípios deontológicos vigentes são suficientes para fundamentar a responsabilização.

A verificação de que o procurador utilizou IA para elaborar a peça sem revisão adequada, resultando em erro material que causou dano à Fazenda Pública, poderia configurar responsabilidade em três planos distintos e cumulativos.

No plano civil, a responsabilidade do procurador perante o ente público que representa fundamenta-se no art. 37, § 6º, da Constituição Federal, que consagra a responsabilidade objetiva do Estado perante terceiros e abre, em seguida, a possibilidade de ação regressiva contra o agente público causador do dano quando este tiver agido com dolo ou culpa. A responsabilidade regressiva pressupõe a demonstração de conduta culposa ou dolosa do procurador — no caso em exame, a negligência consubstanciada na ausência de revisão adequada da peça gerada por IA —, do dano efetivo ao erário e do nexo de causalidade entre ambos. O regime jurídico aplicável a essa responsabilização varia conforme o vínculo funcional do procurador: servidores estatutários federais, estaduais ou municipais sujeitam-se às respectivas leis e aos respectivos regimes jurídicos próprios; empregados públicos, ao regime celetista; e membros de carreiras com estatutos específicos, às normas de regência de sua carreira, sem prejuízo da incidência subsidiária dos estatutos gerais. O denominador comum a todos esses regimes é a exigência constitucional do dolo ou culpa como pressuposto da ação regressiva, o que torna juridicamente relevante, para fins de responsabilização, a demonstração de que o procurador deixou de exercer a supervisão humana que lhe competia sobre o conteúdo gerado pela ferramenta de inteligência artificial.

No plano disciplinar, os procuradores não se sujeitam apenas aos regimes gerais dos servidores públicos, respondendo prioritariamente perante os órgãos correicionais próprios de suas carreiras, com base nas legislações de regência específicas. A diversidade institucional da advocacia pública brasileira implica pluralidade de regimes disciplinares:

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

no âmbito federal, os membros da Advocacia-Geral da União são regidos pela Lei Complementar n.º 73/1993, que estrutura a carreira, define os deveres funcionais e estabelece o regime disciplinar aplicável, com incidência subsidiária das proibições, deveres e sanções previstos no estatuto geral dos servidores federais; no âmbito estadual e distrital, os Procuradores dos Estados e do Distrito Federal submetem-se às respectivas leis orgânicas e estatutos da advocacia pública local, que em regra atribuem às corregedorias das próprias Procuradorias a competência para apurar infrações de seus membros; no âmbito municipal, os Procuradores dos Municípios sujeitam-se aos estatutos funcionais municipais e, onde existam, aos regulamentos disciplinares específicos das Procuradorias; os procuradores das Assembleias Legislativas, Câmaras Municipais e demais órgãos do Poder Legislativo vinculam-se aos regimes disciplinares próprios de cada casa. Em todos esses casos, a edição de política institucional formal de uso de IA confere respaldo normativo adicional à instauração de processo disciplinar em caso de descumprimento das diretrizes estabelecidas, tornando mais objetiva a demonstração da infração funcional. Os tipos disciplinares mais comuns a todos esses estatutos — e potencialmente aplicáveis à hipótese em exame — são a negligência no desempenho das atribuições, o descumprimento de regulamentos internos e a prática de atos incompatíveis com a dignidade do cargo.

É importante notar que a responsabilidade disciplinar pressupõe a demonstração de conduta culposa ou dolosa. O mero uso de IA como auxílio à elaboração de peças processuais, desde que acompanhado de supervisão humana adequada e de verificação das informações geradas, não configura, por si só, qualquer infração. A ilicitude surge quando a ferramenta é utilizada de forma irresponsável, sem a devida diligência profissional, resultando em erro que causou — ou poderia causar — dano ao interesse público.

7.3 Accountability e transparência

Uma questão ainda em aberto na doutrina e na prática é a da obrigação de disclosure — ou seja, se o procurador tem o dever de informar ao juízo e à parte adversa que utilizou sistemas de IA na elaboração da peça processual. Alguns tribunais norte-americanos já editaram regras nesse sentido, exigindo que as partes declarem, em petição, se utilizaram IA generativa na elaboração. No Brasil, o Conselho Nacional de Justiça iniciou o debate sobre o tema, mas ainda não há norma vinculante que imponha esse dever à advocacia pública.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

Do ponto de vista ético, há argumentos relevantes tanto a favor quanto contra a obrigação de disclosure. A favor, argumenta-se que a transparência quanto ao uso de IA permite ao juízo e à parte adversa conhecer o processo de elaboração da peça e exercer maior escrutínio sobre seu conteúdo. Contrarrebuta-se que a divulgação do uso de ferramentas de trabalho não é exigida em relação a outros instrumentos tecnológicos, como editores de texto ou bases de dados jurídicas, e que a responsabilidade do procurador pela peça é integral independentemente de como ela foi elaborada.

A posição mais equilibrada, a nosso ver, é a de que, na ausência de norma que imponha o dever, não há obrigação de informar o uso de IA como ferramenta de apoio. Contudo, quando o uso da IA resultar em erro identificado pelo próprio procurador, a ética profissional impõe a retificação imediata da peça, independentemente de as consequências já terem se produzido.

8 LIMITES ÉTICOS E JURÍDICOS DA DELEGAÇÃO DE ATIVIDADES INTELECTUAIS À IA

8.1 A insubstituibilidade do julgamento profissional

A análise dos potenciais benefícios e riscos da IA na Advocacia de Estado conduz inevitavelmente à questão filosófica mais profunda que subjaz a todo o debate: existem atividades intelectuais inerentes à advocacia que não podem ser delegadas a sistemas automatizados, por razões que não são meramente técnicas — ligadas à atual incapacidade do modelo —, mas, estruturalmente, éticas e jurídicas?

A resposta, a nosso ver, é afirmativa. A advocacia, como profissão jurídica, é fundamentalmente um exercício de julgamento — a capacidade de avaliar, em um contexto específico e irrepetível, qual é a conduta mais adequada à luz dos fatos, do direito e dos valores que regem o ordenamento jurídico. Esse julgamento não é redutível a um processo de reconhecimento de padrões, que é o que os LLMs fazem: envolve uma compreensão genuína das normas jurídicas, uma avaliação ética dos interesses em conflito e uma responsabilidade perante as consequências da decisão.

No contexto da Advocacia de Estado, esse caráter insubstituível do julgamento profissional é ainda mais acentuado. O procurador não representa apenas o interesse

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

imediato do ente público em um litígio: representa o interesse público primário, que inclui a observância da legalidade, a proteção do erário e o respeito aos direitos dos cidadãos. Um sistema de IA, por mais sofisticado que seja, não possui — e não pode possuir — essa dimensão de responsabilidade perante o interesse coletivo.

8.2 Limites éticos da automação

Do ponto de vista ético, a delegação de atividades intelectuais à IA suscita questões que a deontologia jurídica ainda está em processo de enfrentar de forma sistemática. Um primeiro limite claro é o da autoria intelectual: ao assinar uma peça que sabe ter sido elaborada predominantemente por um sistema de IA sem supervisão adequada, o procurador estaria atestando implicitamente a autoria de um trabalho que não é seu — o que pode ser considerado uma forma de desonestidade intelectual incompatível com a deontologia profissional.

Um segundo limite é o do interesse do cliente — no caso da Advocacia de Estado, o interesse público. A utilização de IA sem supervisão adequada expõe o interesse representado a riscos que o procurador seria capaz de evitar: a inclusão de argumentos factualmente incorretos, a omissão de defesas relevantes e a adoção de estratégias processuais inadequadas ao caso concreto. A subordinação do julgamento profissional à saída automatizada do modelo, sem revisão crítica, viola o dever de lealdade ao interesse representado.

Um terceiro limite é o da lealdade processual. O art. 5º do Código de Processo Civil e o art. 77 estabelecem deveres de boa-fé e de lealdade processual aplicáveis às partes e aos seus advogados. A apresentação de argumentos gerados por IA que o procurador não verificou — e que podem ser factualmente incorretos — pode configurar violação desses deveres, especialmente se resultar na apresentação de precedentes inexistentes ou de afirmações fáticas sem respaldo documental.

8.3 Limites jurídicos: o que a norma atual permite e veda

No estado atual da legislação brasileira, não há norma que proíba expressamente o uso de IA na elaboração de peças processuais. O que existe são princípios gerais que,

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

interpretados à luz das características específicas da tecnologia, estabelecem os limites do uso permitido.

São usos claramente permitidos: o emprego de IA como ferramenta de pesquisa e de apoio à elaboração de minutas, desde que o procurador realize supervisão humana qualificada e assuma integral responsabilidade pelo conteúdo da peça final; a utilização de IA para análise e síntese de documentos; a criação de bibliotecas de prompts padronizados para agilizar a produção de manifestações em matérias repetitivas.

São usos claramente vedados ou eticamente questionáveis: a submissão de peças elaboradas por IA sem revisão humana; a inserção de informações sigilosas ou de dados pessoais protegidos pela LGPD em ferramentas de IA não homologadas pelo órgão; a utilização de técnicas de jailbreak para induzir o modelo a produzir conteúdo que as salvaguardas éticas do sistema vedariam; e a atribuição de autoria à IA em peças assinadas pelo procurador sem conhecimento do juízo.

9 DIRETRIZES PARA USO SEGURO POR PROCURADORES

Sintetizando os elementos analisados ao longo do presente estudo, propõem-se as seguintes diretrizes para o uso seguro de sistemas de IA por procuradores de Estado:

1. Princípio da Supervisão Humana Obrigatória: toda peça processual que conte com qualquer participação de sistema de IA em sua elaboração deve ser integralmente revisada pelo procurador responsável antes da assinatura. A revisão deve ser substantiva — não meramente formal — e compreender a verificação de todos os elementos fáticos, jurídicos e jurisprudenciais relevantes.

2. Princípio da Verificação de Fontes: nenhuma referência jurisprudencial, legislativa ou doutrinária gerada por sistema de IA deve ser inserida em peça processual sem verificação prévia nas fontes primárias. A presunção, para fins de segurança institucional, deve ser sempre a de que o modelo pode ter alucinado a referência.

3. Princípio da Proporcionalidade do Uso: a participação da IA na elaboração da peça deve ser proporcional à complexidade e ao impacto do caso. Em causas de natureza rotineira e de baixo valor, o uso de minutas geradas por IA, com revisão simplificada, pode ser adequado. Em causas complexas, de grande impacto ou com potencial de gerar precedentes para outros casos, a participação da IA deve ser mais limitada e a supervisão humana, mais rigorosa.

RCMOS – Revista Científica Multidisciplinar O Saber.
ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

4. Princípio da Confidencialidade: os procuradores devem respeitar rigorosamente as diretrizes institucionais sobre quais informações podem ser inseridas em sistemas de IA. Informações classificadas, dados pessoais protegidos pela LGPD e informações sobre estratégias processuais sensíveis nunca devem ser inseridas em ferramentas não homologadas pelo órgão.

5. Princípio da Responsabilidade Integral: o procurador deve ter plena consciência de que é o único responsável pelo conteúdo da peça assinada, independentemente do grau de participação da IA em sua elaboração. A utilização de IA não atenua nem exclui a responsabilidade profissional, disciplinar ou civil.

6. Princípio da Atualização Contínua: diante da velocidade de evolução da tecnologia de IA, os procuradores devem manter-se atualizados sobre as capacidades, limitações e riscos das ferramentas que utilizam, participando dos programas de capacitação institucional e buscando atualização permanente sobre o tema.

7. Princípio do Reporte de Incidentes: qualquer incidente relacionado ao uso inadequado de IA — identificação de alucinação em peça já apresentada, suspeita de prompt injection, vazamento de informações — deve ser imediatamente reportado ao comitê de governança do órgão, para que as medidas corretivas possam ser adotadas.

CONCLUSÃO

O presente artigo demonstrou que a inteligência artificial generativa representa uma transformação significativa e irreversível no ambiente de trabalho da Advocacia de Estado, com potencial para elevar substancialmente a produtividade dos procuradores nas atividades de elaboração de minutas, pesquisa jurisprudencial e consolidação de argumentos. Contudo, esse potencial somente se realizará de forma benéfica se a tecnologia for utilizada com consciência de seus riscos, dentro de um marco institucional de governança rigorosa e com preservação integral do julgamento profissional do procurador.

As alucinações dos modelos de linguagem, o risco de prompt injection, o vazamento de informações sensíveis e a possibilidade de manipulação de respostas são ameaças reais que demandam respostas técnicas, procedimentais e institucionais coordenadas. Nenhuma dessas ameaças é, por si só, suficiente para justificar a proibição do uso de IA na Advocacia de Estado — tal postura seria tão equivocada quanto a adoção acrítica e irrestrita da

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

tecnologia. O caminho é o da regulamentação responsável, que maximize os benefícios e minimize os riscos.

A engenharia de prompt aplicada à advocacia pública e a criação de bibliotecas institucionais de prompts padronizados representam instrumentos concretos para elevar a qualidade das saídas geradas por IA, ao mesmo tempo em que contribuem para a padronização das manifestações e para a proteção institucional contra os riscos identificados. A governança do uso de IA — com política formal, comitê específico, programa de capacitação e auditoria periódica — é condição necessária para que a tecnologia seja incorporada de forma responsável ao cotidiano dos órgãos de advocacia de Estado.

No plano da responsabilidade, a análise empreendida permite concluir que o procurador é integralmente responsável pelo conteúdo de toda peça que assina, independentemente da participação de sistemas de IA em sua elaboração. A violação dos deveres de diligência e de supervisão, quando resultar em dano ao interesse público, pode configurar responsabilidade disciplinar, funcional e civil. O reconhecimento dessa responsabilidade não deve inibir o uso da tecnologia — deve, ao contrário, incentivar o uso criterioso e tecnicamente fundamentado.

Por fim, cabe observar que a regulamentação do uso de IA na advocacia pública é matéria que transcende os limites de cada órgão e demanda ação coordenada nos planos federal e estadual. A edição de diretrizes pelo Conselho Federal da OAB, pelo Conselho Nacional de Justiça e pelo próprio Governo Federal — com base em estudos técnicos aprofundados e na experiência comparada — é uma medida urgente para que o Brasil desenvolva um marco normativo adequado ao desafio que a inteligência artificial representa para o Estado de Direito.

REFERÊNCIAS

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Senado Federal, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 1 jun. 2026.

BRASIL. Lei n.º 8.112, de 11 de dezembro de 1990. Dispõe sobre o regime jurídico dos servidores públicos civis da União. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/L8112cons.htm. Acesso em: 1 jun. 2026.

BRASIL. Lei n.º 8.906, de 4 de julho de 1994. Dispõe sobre o Estatuto da Advocacia e a Ordem dos Advogados do Brasil (OAB). Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8906.htm. Acesso em: 1 jun. 2026.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

BRASIL. Lei n.º 13.105, de 16 de março de 2015. Código de Processo Civil. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 1 jun. 2026.

BRASIL. Lei n.º 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 1 jun. 2026.

BRASIL. Conselho Nacional de Justiça. Resolução CNJ n.º 332, de 21 de agosto de 2020. Dispõe sobre ética, transparência e governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 1 jun. 2026.

BRASIL. Conselho Nacional de Justiça. Resolução CNJ n.º 433, de 2023. Estabelece diretrizes para o uso de inteligência artificial pelos tribunais. Disponível em: https://atos.cnj.jus.br/. Acesso em: 1 jun. 2026.

BRASIL. Tribunal Superior Eleitoral. Ação de Investigação Judicial Eleitoral n.º 0600814-85.2022.6.00.0000. Corregedor-Geral da Justiça Eleitoral: Min. Benedito Gonçalves. Decisão de 14 de abril de 2023. Disponível em: https://www.conjur.com.br/2023-abr-18/tse-multa-advogado-peticao-baseada-conversa-chatgpt/. Acesso em: 4 jun. 2026.

BRASIL. Tribunal de Justiça de Santa Catarina. 6ª Câmara Cível. Multa decorrente de jurisprudência falsa gerada por inteligência artificial em recurso de reintegração de posse. Decisão de 19 de fevereiro de 2025. Disponível em: https://www.tjsc.jus.br/web/imprensa/-/tjsc-multa-autor-de-recurso-por-jurisprudencia-falsa-gerada-por-ia. Acesso em: 4 jun. 2026.

BRASIL. Tribunal de Justiça de Santa Catarina. 5ª Câmara Criminal. Advertência ao advogado por habeas corpus redigido por IA, com jurisprudência inexistente. Decisão de fevereiro de 2025. In: MIGALHAS. TJ/SC adverte advogado por HC elaborado por IA com jurisprudência falsa. São Paulo, 10 fev. 2025. Disponível em: https://www.migalhas.com.br/quentes/424313/tj-sc-adverte-advogado-por-hc-feito-por-ia-com-jurisprudencia-falsa. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 7ª Região. 3ª Turma. Processo n.º 0000702-38.2024.5.07.0016. Multa por litigância de má-fé e ofício à OAB-CE por jurisprudência fictícia com indícios de geração por IA. Decisão de junho de 2025. In: MIGALHAS. TRT-7: suspeita de jurisprudência gerada por IA e multa ao advogado. São Paulo, 12 jun. 2025. Disponível em: https://www.migalhas.com.br/quentes/432475/trt-7-suspeita-de-jurisprudencia-gerada-por-ia-e-multa-advogado. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 12ª Região. Vara do Trabalho de Concórdia. Ação trabalhista com petição inicial contendo decisões, citação doutrinária e magistrado inexistentes gerados por IA. Juiz Daniel Carvalho Martins. Decisão de outubro de 2025. Disponível em: https://portal.trt12.jus.br/noticias/autora-de-acao-e-multada-apos-advogada-inventar-jurisprudencia-e-desembargador. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 2ª Região. Processo n.º 1001128-84.2024.5.02.0044. Multa por má-fé e ofício à OAB-SP pelo uso de jurisprudência fictícia gerada por IA. Relator: Juiz convocado Fernando César Teixeira França. Decisão de fevereiro de 2026. In: CONJUR. O uso de jurisprudência falsa gerada por IA acarreta multa e ofício à OAB. São Paulo, 16 fev. 2026. Disponível em: https://www.conjur.com.br/2026-fev-16/uso-de-jurisprudencia-criada-por-ia-gera-multa-por-ma-fe-e-oficio-a-oab/. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 18ª Região. Ação Trabalhista – Rito Sumaríssimo n.º 0001204-10.2025.5.18.0121. Multa por má-fé e ofício à OAB-GO pelo uso

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

de jurisprudência falsa atribuída à IA. Decisão de maio de 2026. In: ROTA JURÍDICA. Uso de jurisprudência atribuída à IA gera multa por má-fé e ofício à OAB de Goiás. 2026. Disponível em: https://www.rotajuridica.com.br/uso-de-jurisprudencia-atribuida-a-ia-gera-multa-por-ma-fe-e-oficio-a-oab-de-goias/. Acesso em: 4 jun. 2026.

BRASIL. Superior Tribunal de Justiça. Nota oficial: tentativas de uso de *prompt injection* no STJ serão investigadas. Brasília: STJ, 20 de maio de 2026. Disponível em: https://www.stj.jus.br/sites/portalp/paginas/comunicacao/noticias/2026/20052026-tentativas-de-uso-de-prompt-injection-no-stj-serao-investigadas.aspx. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 8ª Região. 3ª Vara do Trabalho de Parauapebas. Sentença. Processo ATOrd n.º 0001062-55.2025.5.08.0130. Juiz substituto: Luiz Carlos de Araújo Santos Júnior. Decisão de 12 de maio de 2026. In: MIGALHAS. *Prompt injection* oculto na petição inicial: o caso de Parauapebas. São Paulo, maio de 2026. Disponível em: https://www.migalhas.com.br/depeso/455925/prompt-injection-oculto-em-peticao-inicial-o-caso-de-parauapebas. Acesso em: 4 jun. 2026.

BRASIL. Tribunal Regional do Trabalho da 8ª Região. 3ª Vara do Trabalho de Parauapebas. Sentença. Processo ATOrd n.º 0001062-55.2025.5.08.0130. In: MIGALHAS. *Prompt injection* na petição inicial: o caso de Parauapebas — Migalhas de Responsabilidade Civil. São Paulo, maio de 2026. Disponível em: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/456362/prompt-injection-em-peticao-inicial-o-caso-de-parauapebas. Acesso em: 4 jun. 2026.

BRASIL. Conselho Nacional de Justiça. Resolução n.º 615, de 2025. Dispõe sobre o uso de inteligência artificial generativa no Poder Judiciário. Brasília: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 4 jun. 2026.

BROWN, Tom et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, v. 33, p. 1877-1901, 2020. Disponível em: https://arxiv.org/abs/2005.14165. Acesso em: 1 jun. 2026.

CHUI, Michael et al. The economic potential of generative AI: The next productivity frontier. McKinsey Global Institute, jun. 2023. Disponível em: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier. Acesso em: 1 jun. 2026.

CONJUR. Tentativas de uso de *prompt injection* no STJ serão investigadas. São Paulo, 20 de maio de 2026. Disponível em: https://www.conjur.com.br/2026-mai-20/tentativas-de-uso-de-prompt-injection-no-stj-serao-investigadas/. Acesso em: 4 jun. 2026.

CONSTITUTIONAL AI PRINCIPLES. AI safety and alignment: key frameworks. In: ANTHROPIC Research Blog. San Francisco: Anthropic, 2023. Disponível em: https://www.anthropic.com/research. Acesso em: 1 jun. 2026.

DAFOE, Allan. AI Governance: A Research Agenda. Future of Humanity Institute, University of Oxford, 2018. Disponível em: https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf. Acesso em: 1 jun. 2026.

DATA PRIVACY BRASIL. *Prompt injection* no Judiciário expõe o custo da IA sem governança. São Paulo, jun. 2026. Disponível em: https://www.dataprivacybr.org/prompt-injection-no-judiciario-expoe-o-custo-da-ia-sem-governanca/. Acesso em: 4 jun. 2026.

HALLUCINATION LEADERBOARD. Vectara. Disponível em: https://github.com/vectara/hallucination-leaderboard. Acesso em: 1 jun. 2026.

JI, Ziwei et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, v. 55, n. 12, p. 1-38, 2023. Disponível em: https://arxiv.org/abs/2202.03629. Acesso em: 1 jun. 2026.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

JUSDOCS. *Prompt injection* no processo: multa, OAB e inquérito do STJ. 2026. Disponível em: https://jusdocs.com/blog/prompt-injection-processo-multa-oab-stj-inquerito-ia. Acesso em: 4 jun. 2026.

KASPERSKY. O que é *prompt injection*? Como a IA pode ser manipulada e o que você pode fazer. 2026. Disponível em: https://www.kaspersky.com.br/resource-center/threats/prompt-injection. Acesso em: 4 jun. 2026.

KATZ, Daniel Martin; BOMMARITO, Michael; BLACKMAN, Josh. GPT-4 Passes the Bar Exam. Social Science Research Network, 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233. Acesso em: 1 jun. 2026.

MIGALHAS. *Prompt injection* é "caso de polícia", alerta ministro Salomão. São Paulo, jun. 2026. Disponível em: https://www.migalhas.com.br/quentes/457223/prompt-injection-e-caso-de-policia-alerta-ministro-salomao. Acesso em: 4 jun. 2026.

MIGALHAS. *Prompt injection*: STJ apura tentativa de manipulação de IA em petições. São Paulo, maio de 2026. Disponível em: https://www.migalhas.com.br/quentes/456424/prompt-injection-stj-apura-tentativa-de-manipulacao-de-ia-em-peticoes. Acesso em: 4 jun. 2026.

MIGALHAS. *Prompt injection*: a IA criou o vetor, não a má conduta. São Paulo, maio de 2026. Disponível em: https://www.migalhas.com.br/amp/depeso/456010/prompt-injection-a-ia-criou-o-vetor-nao-a-ma-conduta. Acesso em: 4 jun. 2026.

MCNAUGHTON, David. Prompt Injection Attacks Against LLM-Integrated Applications. arxiv, 2023. Disponível em: https://arxiv.org/abs/2302.12173. Acesso em: 1 jun. 2026.

OPENAI. GPT-4 Technical Report. arxiv, 2023. Disponível em: https://arxiv.org/abs/2303.08774. Acesso em: 1 jun. 2026.

ORDEM DOS ADVOGADOS DO BRASIL. Código de Ética e Disciplina da OAB. Aprovado pela Resolução n.º 02/2015. Brasília: OAB, 2015. Disponível em: https://www.oab.org.br/content/pdf/legislacaooab/codigodeetica.pdf. Acesso em: 1 jun. 2026.

ORDEM DOS ADVOGADOS DO BRASIL. Conselho Federal. Recomendação n.º 1, de 2024. Dispõe sobre o uso de inteligência artificial generativa na prática jurídica. Brasília: OAB, 2024.

OWASP — OPEN WORLDWIDE APPLICATION SECURITY PROJECT. *OWASP Top 10 for Large Language Model Applications 2025*. Disponível em: https://owasp.org/www-project-top-10-for-large-language-model-applications/. Acesso em: 4 jun. 2026.

PEREZ, Ethan; RIBEIRO, Marco Tulio. Ignore Previous Prompt: Attack Techniques for Language Models. arxiv, 2022. Disponível em: https://arxiv.org/abs/2211.09527. Acesso em: 1 jun. 2026.

ROMM, Tony. A lawyer cited ChatGPT in court. The judge was not amused. The Washington Post, 25 maio 2023. Disponível em: https://www.washingtonpost.com/technology/2023/05/27/chatgpt-ai-lawsuit-avianca-airlines/. Acesso em: 1 jun. 2026.

VASWANI, Ashish et al. Attention Is All You Need. Advances in Neural Information Processing Systems, v. 30, 2017. Disponível em: https://arxiv.org/abs/1706.03762. Acesso em: 1 jun. 2026.

WEI, Jason et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. Advances in Neural Information Processing Systems, v. 35, 2022. Disponível em: https://arxiv.org/abs/2201.11903. Acesso em: 1 jun. 2026.

Ano VII, v.1 2026 | submissão: 04/06/2026 | aceito: 07/06/2026 | publicação: 10/06/2026

WHITE, Jules et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arxiv, 2023. Disponível em: https://arxiv.org/abs/2302.11382. Acesso em: 1 jun. 2026.

WORLD ECONOMIC FORUM. Applying AI in Legal Contexts: Opportunities, Risks and Governance. Geneva: WEF, 2024. Disponível em: https://www.weforum.org/reports/. Acesso em: 1 jun. 2026.