Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Inteligencia artificial y defensa del Estado: parámetros para un uso seguro en la redacción de documentos legales.
Inteligencia artificial y asesoría jurídica estatal: parámetros para un uso seguro en la redacción de documentos procesales.
Inteligencia artificial y defensa del Estado: parámetros para un uso seguro en la elaboración de documentos procesales.

Leandro Mendes Neris
Licenciado en Derecho
Institución: Universidad Federal de Amazonas
Correo electrónico: leandromendesneris@gmail.com

RESUMEN: Este artículo analiza, desde perspectivas legales, éticas y técnicas, la incorporación de sistemas de inteligencia artificial generativa —en particular, modelos lingüísticos a gran escala (MLG)— en las actividades de los fiscales estatales, especialmente en la redacción de documentos procesales, dictámenes, apelaciones y peticiones. El estudio examina los riesgos específicos de esta tecnología en el contexto jurídico-institucional, incluyendo el fenómeno de las alucinaciones del modelo lingüístico, la vulnerabilidad a los ataques de inyección de mensajes, el riesgo de filtraciones de información confidencial y la manipulación de las respuestas generadas por la IA. Por otro lado, se analizan los beneficios potenciales en términos de productividad, investigación jurisprudencial y estandarización.

El artículo propone parámetros prácticos para un uso seguro, incluyendo técnicas de ingeniería de mensajes rápidos aplicadas a la defensa pública, gobernanza institucional, bibliotecas de mensajes estandarizadas y criterios de responsabilidad disciplinaria para los fiscales. Concluye que la IA puede ser una herramienta poderosa para la defensa pública, siempre que su uso esté regulado por supervisión humana cualificada, protocolos institucionales estrictos y...
Conciencia de los límites éticos y legales de delegar actividades intelectuales a sistemas automatizados.

PALABRAS CLAVE: Inteligencia artificial. Defensa estatal. Inyección inmediata.
Alucinaciones del máster en Derecho. Ingeniería rápida. Responsabilidad del abogado. Gobernanza institucional. Seguridad de la información.

RESUMEN: Este artículo analiza, desde perspectivas legales, éticas y técnicas, la incorporación de sistemas de inteligencia artificial generativa —en particular, los Modelos de Lenguaje a Gran Escala (MLE)— en las actividades de la Asesoría Jurídica del Estado, especialmente en la redacción de escritos procesales, dictámenes jurídicos, apelaciones y peticiones. El estudio examina los riesgos específicos de esta tecnología en el contexto jurídico-institucional, incluyendo la manipulación de los modelos de lenguaje, la vulnerabilidad a ataques de inyección de mensajes, el riesgo de fuga de información sensible y la manipulación de las respuestas generadas por la IA. Por otro lado, también se analizan los beneficios potenciales en términos de productividad, investigación jurisprudencial y estandarización. El artículo propone parámetros prácticos para un uso seguro, incluyendo técnicas de ingeniería de mensajes aplicadas a la asesoría jurídica pública, gobernanza institucional, bibliotecas de mensajes estandarizadas y criterios para la responsabilidad disciplinaria del abogado. Concluye que la IA puede ser una herramienta poderosa para la Asesoría Jurídica del Estado, siempre que su uso esté regulado por supervisión humana cualificada, protocolos institucionales estrictos y la conciencia de los límites éticos y legales de delegar actividades intelectuales a sistemas automatizados.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

PALABRAS CLAVE: Inteligencia artificial. Asesoría jurídica estatal. Inyección inmediata. Alucinaciones de LLM. Ingeniería inmediata. Responsabilidad del abogado. Gobernanza institucional.
Seguridad de la información.

1. INTRODUCCIÓN

La revolución tecnológica iniciada por la difusión de los modelos de lenguaje.
Los modelos de lenguaje a gran escala (LLM) posiblemente representan los más
Desde la popularización de internet, se ha producido una profunda transformación en las prácticas intelectuales.
En el ámbito jurídico, el uso de herramientas basadas en inteligencia artificial generativa —
como ChatGPT de OpenAI; Claude de Anthropic; Gemini de Google; y Copilot,
En el caso de Microsoft, ha dejado de ser una perspectiva futura para convertirse en una realidad presente.
incorporado a la rutina diaria de los bufetes de abogados y departamentos jurídicos.
primero en el ámbito corporativo y, progresivamente, en carreras jurídicas en el sector público.

Defensa del Estado — entendida, en el sistema jurídico brasileño, como
La Oficina del Fiscal General, a través de las Oficinas de los Fiscales de Distrito Estatales y Federales y la
Las oficinas de los abogados municipales enfrentan desafíos particulares en este escenario. Por un lado,
la creciente demanda de eficiencia, optimización de recursos y velocidad de respuesta en
En contextos con un alto volumen de procesos, la inteligencia artificial se convierte en una solución.
atractivo. Por otro lado, la peculiar naturaleza de la función de representación judicial y consultiva de
El Estado, guiado por la inalienabilidad del interés público, por la estricta observancia de
La legalidad y la responsabilidad institucional del fiscal imponen límites que no
Existen —o existen en menor medida— en la práctica privada.

El uso de la IA en la redacción de documentos legales plantea interrogantes que
Invaden diferentes ámbitos del conocimiento jurídico. Desde un punto de vista técnico, es
Es necesario comprender los mecanismos de funcionamiento de los LLM y sus puntos débiles.
y sus vulnerabilidades. Desde un punto de vista ético y profesional, surge el problema de
delegación de actividades intelectuales inherentes al ejercicio de la abogacía a sistemas
procesos automatizados, con implicaciones para el deber de diligencia, la lealtad procedimental y...
La responsabilidad del abogado. Desde un punto de vista institucional, surgen interrogantes sobre...
La gobernanza, la seguridad de la información y la estandarización de los procedimientos requieren
Reglamentos internos de los organismos de asistencia jurídica del Estado.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

El objetivo de este artículo es analizar, de forma sistemática y práctica, la
Parámetros para el uso seguro de sistemas de inteligencia artificial en la profesión jurídica.
Estado. Con ese fin, el estudio se estructura de la siguiente manera: inicialmente, un
Se presenta una descripción general de los LLM y su funcionamiento; luego, se examina el fenómeno de...
Se analizan las alucinaciones y sus repercusiones legales, el concepto de inyección inmediata y otros temas relacionados.
Se identifican los vectores de riesgo; se proponen medidas de mitigación y directrices de gobernanza; y se analiza lo siguiente.
Ingeniería eficaz aplicada a la defensa de los intereses públicos; y, finalmente, se analiza la responsabilidad.
Se iniciarán acciones disciplinarias y legales contra el fiscal por utilizar contenido generado por inteligencia artificial.

2. Modelos de lenguaje a gran escala y contexto.
LEGAL

2.1 Funcionamiento básico de los LLM

Los modelos de lenguaje a gran escala son sistemas de inteligencia artificial.
entrenados con grandes volúmenes de texto, con el objetivo de predecir, de forma probabilística,
¿Qué token (unidad de texto) debe seguir al anterior en una secuencia? Comenzando desde
Arquitecturas basadas en transformadores, desarrolladas a partir del artículo fundamental de Vaswani.
et al. (2017), estos modelos aprendieron patrones lingüísticos suficientemente ricos para
Generar textos coherentes, responder preguntas complejas, traducir idiomas, redactar documentos.
y, si resulta de interés, la redacción de documentos legales.

Es esencial entender que un LLM no "conoce" correctamente en el sentido de que un
Un experto legal sabe: el modelo no tiene representaciones internas de normas, no consulta bases de datos de
datos en tiempo real (excepto cuando se cuenta con herramientas externas específicas) y no
Razona deductivamente a partir de premisas establecidas. Lo que hace el modelo es:
Esencialmente, reconocer patrones estadísticos y generar texto que, dados los datos de entrada,
recibido (indicaciones), se asemeja a lo que, en su conjunto de entrenamiento, seguiría a eso
contexto. La apariencia de un razonamiento jurídico riguroso es, en gran medida, una consecuencia.
a partir de la estructura del texto legal incorporado a la formación, y no de un proceso deductivo.
genuino.

Esta distinción es fundamental para comprender los riesgos asociados con el uso de
Los másteres en Derecho (LLM) aplicados a la práctica jurídica explican por qué este modelo puede generar argumentos jurídicamente sólidos.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Impecable en la forma, pero incorrecto en el contenido — citando
precedentes inexistentes, atribuir resúmenes falsos a fallos reales o afirmar que el estado
reglamentos relativos a leyes que ya han sido derogadas.

2.2 El LLM como herramienta de apoyo y sus limitaciones

Desde el punto de vista de la productividad, los LLM ofrecen contribuciones reales y documentado en tres áreas principales. En primer lugar, en acelerar la producción de Minutos: Un fiscal que domine las técnicas de ingeniería rápida puede obtener un Redacción estructurada de una defensa, respuesta a objeciones u opinión en una fracción del tiempo lo que requeriría producirlo desde cero, dedicando energía intelectual a la supervisión y el ajuste. y a la mejora del contenido generado. En segundo lugar, en la investigación de jurisprudencia, el Los modelos con acceso a internet o equipados con bases de datos legales integradas pueden identificar precedentes relevantes, sintetizar argumentos legales y trazar la evolución de la jurisprudencia con mayor velocidad que la investigación manual. En tercer lugar, en la estandarización del lenguaje. y en cumplimiento formal, reduciendo los errores de procedimiento de carácter técnico.

Sin embargo, estos beneficios no pueden ocultar las limitaciones estructurales críticas. El conocimiento de los LLM es estático, limitado a la fecha límite de la formación y frecuentemente Obsoleto en relación con los cambios legislativos y jurisprudenciales más recientes. La alucinación —la producción segura de información falsa— es un fenómeno intrínseco a El funcionamiento de estos modelos no es un fallo que pueda corregirse con simples ajustes. Y el La dependencia excesiva de los resultados del modelo, sin supervisión humana cualificada, puede transformando la IA de una herramienta de apoyo en un vehículo para errores materiales, con graves consecuencias procesales.

3. Alucinaciones de los modelos lingüísticos y su impacto en el desempeño del Tesoro Público

3.1 Concepto y tipología de las alucinaciones

El término alucinación, tal como se aplica a los LLM, se refiere a la generación de información que no Corresponden a la realidad fáctica, pero el modelo las presenta con el mismo nivel de... Confianza y fluidez que brindan información precisa. A diferencia de una errata.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

o una inconsistencia lógica fácilmente perceptible, las alucinaciones de los LLM
A menudo se presentan en forma de texto bien estructurado, internamente coherente y
aparentemente plausible, lo que hace que su detección sea particularmente difícil para aquellos
No tiene un conocimiento profundo del tema.

En un contexto jurídico, las alucinaciones adquieren tipologías específicas de particular importancia.
El primer ejemplo, y el más documentado, es la falsificación de precedentes judiciales: el modelo.
cita resoluciones con numeración plausible, relator, fecha y resumen, que simplemente no
Sí, existen. La segunda es la distorsión de precedentes reales: el modelo cita un juicio que existe,
pero le asigna un resumen o tesis diferente a la del original. El tercero es el
Citar disposiciones legales derogadas como si aún estuvieran vigentes, o afirmar estados de derecho.
regulaciones que ya han sido enmendadas. La cuarta es la atribución errónea de posiciones doctrinales a
autores, citando obras inexistentes o atribuyendo una tesis a un jurista en particular que no es cierta.
su.

3.2 Casos documentados de alucinaciones en la profesión jurídica

El estudio de caso más difundido sobre el impacto de las alucinaciones provocadas por la IA en la práctica.
El argumento legal fue presentado por el abogado Steven Schwartz en Estados Unidos, quien interpuso una petición en 2023.
ante el Tribunal Federal del Distrito Sur de Nueva York, citando seis precedentes inexistentes,
todo generado por ChatGPT. Cuando el juez Kevin Castel lo convocó para probar la
A pesar de la existencia de los fallos, el abogado no lo hizo, sufriendo sanciones disciplinarias y
condena y multa impuestas. El episodio generó amplias repercusiones y motivó la
Elaboración de directrices para el uso de la inteligencia artificial por parte de los tribunales de todo el mundo.

En Brasil, contrariamente a lo que se podría suponer, ya existe un conjunto significativo y
creciente conjunto de precedentes sobre el tema, con decisiones sancionadoras emitidas por la mayoría
Diversos tribunales de todo el país. El primer caso de relevancia institucional data de abril de 2023:
El entonces Inspector General de Justicia Electoral, el Ministro Benedito Gonçalves, del TSE, solicitó
Se impuso una multa de R$ 2.604 a un abogado por litigar de mala fe al presentar una demanda.
Petición redactada íntegramente por ChatGPT, en la que pretendía participar como amicus curiae.
en una demanda electoral que involucra al expresidente Jair Bolsonaro, un proceso en el que él no
no representaba a ninguna de las partes. El Ministro fue enfático al señalar que el abogado había
Se presentó a juicio "una fábula, resultado de una conversación con una inteligencia artificial".

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

En febrero de 2025, el TJSC (Tribunal de Justicia de Santa Catarina) consolidó un estándar de toma de decisiones más severo en dos...
casos distintos y emblemáticos. En el primero, la 6ª Sala Civil multó a un abogado en
El 10% del valor del caso después de que admitiera haber utilizado ChatGPT para preparar la apelación.
en una acción de recuperación de posesión, un recurso legal en el que se utilizan ambas citas de jurisprudencia.
En cuanto al grado en que las referencias a obras jurídicas eran incorrectas o totalmente ficticias. Además de
Como sanción económica, el tribunal ordenó que el caso fuera denunciado ante la OAB/SC (Colegio de Abogados de Santa Catarina) y...
Se remitió una copia de la apelación para revisión disciplinaria. El abogado alegó que el error
Esto ocurrió debido al "uso inadvertido" de ChatGPT. En el segundo caso en Santa Catarina, la Quinta Cámara
Los investigadores criminales advirtieron formalmente al abogado que utilizó inteligencia artificial para redactar una petición de hábeas corpus.
No existe precedente, ya que el ponente afirmó que la conducta constituía un acto de mala fe.
y la falta de respeto hacia el tribunal y que los precedentes presentados "fueron creados para inducir la
"El juez cometió un error".

En los tribunales laborales, el fenómeno ha adquirido proporciones aún más preocupantes.
En junio de 2025, el TRT-7 (Ceará) identificó, en un recurso de apelación, jurisprudencia.
manipulado con pasajes falsificados o inexistentes, incluidos pasajes insertados incorrectamente.
asignado al TST que trataba asuntos completamente diferentes: uno de los resúmenes,
Presentado como si tratara sobre acoso laboral, en realidad se refería a...
"PJ-ización" — con evidencia de haber sido generada por inteligencia artificial. El tribunal impuso una multa.
2% del valor del caso y ordenó que se enviara una carta oficial a la OAB-CE (Colegio Brasileño de Abogados de Ceará) para la investigación de
Posible infracción disciplinaria. En octubre de 2025, el TRT-12 (Santa Catarina) juzgó el caso.
en la que la petición inicial contenía decisiones, citas doctrinales e incluso el nombre de un magistrado.
inexistentes, todos los elementos aparentemente generados por inteligencia artificial. El juez
Daniel Carvalho Martins afirmó que "tales hallazgos van más allá de un simple error administrativo" y
que la pieza fue producida "sin ninguna verificación humana, lo que constituye un acto"
"Documento procesal inexistente". Se ordenó a la parte pagar una multa de R$ 3.700.

En febrero de 2026, el TRT-2 (São Paulo) se enfrentó a una situación que expuso una
Una estrategia de evasión recurrente en la práctica forense: cuando se le pregunta sobre el uso de
jurisprudencia ficticia, el abogado atribuyó la responsabilidad al "cuerpo de pasantes de
oficina", un argumento que el tribunal rechazó vehementemente, afirmando que "la reclamación
"En los tribunales, se trata de un acto exclusivo del abogado, quien también es responsable de su contenido."
El magistrado destacó expresamente que la conducta violó la Recomendación No. 1/2024 de
El Consejo Federal del Colegio de Abogados de Brasil, que exige supervisión humana en el uso de herramientas generativas,

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

y que hubo "creación deliberada de precedentes para reforzar el argumento de la defensa, comprometiendo la seguridad jurídica". Es precisamente este instrumento normativo — Recomendación n.º 1/2024 de El Consejo Federal del Colegio de Abogados de Brasil —que ha sido invocado como un parámetro objetivo en decisiones más recientes. Según la regla, el abogado está obligado a tener "comprensión". limitaciones apropiadas, verificación rigurosa de la información, transparencia para los clientes y otros interlocutores, y se prohíbe la delegación de actos exclusivos de la profesión sin "Supervisión cualificada". La consolidación de esta cadena de precedentes, que abarca el TSE, El TJSC y múltiples TRT demuestran que el problema no es ni incipiente ni episódico: Entre 2025 y 2026, los tribunales brasileños registraron varios casos de abogados multados por Presentar peticiones con jurisprudencia falsa generada por IA, estableciendo un patrón. Esta alarmante situación ya ha llegado a conocimiento de las oficinas del defensor del pueblo del estado.

3.3 Impactos específicos en el desempeño del Tesoro Público

Para la Fiscalía del Estado, las repercusiones de las alucinaciones son significativas. Esto es especialmente grave por razones que se suman a las que afectan al ejercicio privado de la abogacía. En primer lugar, el fiscal representa al Estado, titular de un interés público inalienable. —, de modo que un error material resultante de una alucinación de IA pueda comprometer no No se trata solo de una causa individual, sino que también puede sentar precedentes negativos para la entidad pública. a lo largo de una línea de jurisprudencia. En segundo lugar, la responsabilidad funcional de La responsabilidad de un fiscal público es más intensa que la de un abogado privado: errores que, en la práctica privada, En el ámbito de la defensa de intereses públicos, estas acciones solo conllevarían responsabilidades contractuales y disciplinarias. Determinar la mala conducta en un cargo público, con las consiguientes consecuencias administrativas.

En tercer lugar, la naturaleza de las disputas que involucran al Tesoro Público — frecuentemente relacionado con impuestos, seguridad social, pagos ordenados por la corte, servicios públicos y políticas públicas: implica que un error en una sola petición puede tener un efecto multiplicador. sobre cientos o miles de casos similares procesados por la misma agencia. Una tesis Con deficiencias legales, incorporado a un borrador generado por IA y no sometido a una revisión rigurosa. Si se adopta como modelo institucional, puede difundirse sistemáticamente.

En cuarto lugar, existe el riesgo de que la presión por la productividad, a menudo alto en organismos de defensa pública debido al desequilibrio entre el volumen de

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

los procesos y el número de fiscales fomentan una relación de confianza excesiva en
Salir del modelo reduce la necesidad de supervisión humana para detectar alucinaciones.

4 RIESGOS DE SEGURIDAD INSTITUCIONAL: INYECCIÓN INMEDIATA, FUGAS Y MANEJO

4.1 El concepto de inyección inmediata

El concepto de inyección rápida: La inyección rápida es una categoría de ataque que explora la incapacidad estructural de los LLM para distinguir de manera confiable entre instrucciones proporcionadas por el operador del sistema (mensaje del sistema) y el contenido producido por usuarios o fuentes externas que el modelo procesa. El concepto fue formalmente identificado en 2022 y desde entonces ha recibido una atención cada vez mayor por parte de la comunidad de Seguridad. Para 2025, OWASP, una autoridad mundial en seguridad de aplicaciones, Situó la inyección inmediata como la principal amenaza en su clasificación de riesgos para los sistemas. de IA generativa (OWASP Top 10 para LLM 2025). El ataque consiste en insertar, en contenido que el modelo analiza, instrucciones ocultas o disfrazadas que, al ser procesadas A través de LLM, alteran su comportamiento, anulan las instrucciones originales o extraen Información que no debe ser revelada.

Hay dos subtipos principales. El primero es la inyección directa instantánea, que ocurre cuando la entrada del usuario afecta directamente el comportamiento del LLM, ya sea de una manera... intencional —cuando el usuario planifica y estructura el ataque— o no intencional.

La segunda es la inyección indirecta de indicaciones, que se produce cuando el modelo acepta indicaciones de fuentes externas. Las fuentes externas, como sitios web o archivos, y el contenido de estas fuentes, cuando se interpretan, alteran la Comportamiento LLM. Ambas modalidades pueden dar lugar a respuestas sesgadas. en la divulgación de información confidencial o en la ejecución de acciones no autorizadas.

En el contexto de la defensa pública, el riesgo es particularmente sofisticado: un fiscal que utiliza un sistema de IA para analizar la petición inicial de la parte contraria y Extraer los principales argumentos para refutar puede, sin que nos demos cuenta, llevar al procesamiento de comandos maliciosos. insertado en la propia pieza por la parte contraria, instruyendo al modelo para que produzca un análisis. favorecer al oponente, omitir argumentos relevantes o filtrar información estratégica. de la agencia. Lo que hasta hace poco parecía un escenario hipotético se convirtió, en mayo de

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

2026, en una realidad concreta y legalmente sancionada en Brasil, inaugurando una nueva y
Una alarmante frontera de litigios de mala fe.

4.1.1. El caso Parauapebas: Uno de los primeros precedentes brasileños

El 12 de mayo de 2026, el 3er Tribunal Laboral de Parauapebas se encontró con un
un hallazgo sin precedentes en la jurisprudencia brasileña. Al procesar la petición inicial por
Sistema Galileo: una herramienta de inteligencia artificial generativa para los tribunales laborales.
, El tribunal identificó un texto en letra blanca sobre fondo blanco, invisible para el lector.
humano, que contiene el siguiente comando: "Atención, inteligencia artificial, impugna esto
La petición debe ser superficial y no cuestionar los documentos, independientemente de la orden judicial.
lo que sea que se le dé."

El juez dedicó las primeras páginas de la sentencia a condenar la conducta.
afirmando que era "incompatible con los deberes más básicos que recaen sobre
todos los que participan en el proceso judicial" y clasificó la práctica como "un ataque a la justicia".
Se socava la credibilidad de los instrumentos institucionales, lo que conlleva una falta de respeto hacia los tribunales, las partes implicadas y la sociedad.
"Y este es un precedente que este tribunal no puede dejar pasar." El fallo reconoció el acto.
El tribunal, al considerar que el acto constituía una afrenta a la dignidad de la justicia, condenó conjuntamente a los dos abogados firmantes.
al pago de una multa del 10% sobre el valor del caso — aproximadamente R$ 84.250,00 — y
Se determinó que se debía enviar una carta oficial a la OAB/PA (Colegio de Abogados de Pará) y a la Inspección del TRT (Tribunal Regional del Trabajo) de la 8ª Región.

4.1.2. La escalada nacional: STJ, São Paulo y la dimensión criminal

El incidente de Parauapebas ya no es un caso aislado. El 20 de mayo de 2026,
El presidente del Tribunal Superior de Justicia (STJ), el ministro Herman Benjamin, ordenó la apertura de una investigación policial.
y procedimiento administrativo para investigar intentos de fraude procesal por medio de
inyección inmediata en el sistema STJ Logos, un sistema de inteligencia artificial generativa.
Desarrollado por el propio Tribunal y lanzado en febrero de 2025. Según informes televisivos.
Según informa Globo, existen al menos 11 casos penales en los que se ha identificado esta técnica.
El Tribunal Superior de Justicia (STJ) determinó que los intentos de utilizar el mecanismo, aunque neutralizados, no se llevaron a cabo.
a través de las capas de seguridad del sistema, quedarían certificados en los registros, que
permitió la imposición de sanciones procesales por parte de los ministros de la Corte. El vicepresidente de

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

El ministro de STJ, Luis Felipe Salomão, fue aún más enfático al describir la práctica como
"Caso policial", que establece que "cuando la IA se utiliza indebidamente, de mala fe, lo cual va más allá del alcance de..."
Se trata de una cuestión puramente ética; es un asunto que compete a la policía y debe investigarse tanto a nivel individual como policial.
"tanto administrativos como penales."

El 21 de mayo, el fenómeno llegó a São Paulo, con la identificación de rápidamente
Inyección en petición contra el banco en el tribunal civil de la capital, confirmando que el asunto ha quedado pendiente.
Pasó de ser algo anecdótico a adquirir una dimensión sistémica a nivel nacional.

4.1.3. Implicaciones para la defensa de intereses por parte del Estado

La técnica está clasificada por OWASP como la vulnerabilidad número uno en su...
Lista de riesgos para aplicaciones de lenguaje basadas en modelos. Para el Departamento Jurídico de
En el estado, el factor de riesgo es doble. En el sentido activo, la agencia puede ser víctima de una acción rápida.
La inyección indirecta se produce cuando sus abogados presentan documentos de la parte contraria para su revisión.
Sistemas de IA: una situación en la que los comandos ocultos en estos componentes pueden comprometer la
La calidad del análisis o, en sistemas mal configurados, la extracción de información confidencial.
sentido pasivo, sistemas de IA utilizados por los propios organismos de defensa pública —
para la evaluación de procedimientos, el análisis de riesgos y la redacción de actas: estos son objetivos de
ataques externos.

4.2 Fuga de información sensible

El uso de herramientas de IA en el contexto de la defensa del Estado implica:
necesariamente, el procesamiento de información altamente sensible: datos sobre litigios.
Estrategias de defensa y ataque en curso del Tesoro Público, opiniones internas no
Información publicada sobre contribuyentes en procedimientos tributarios, datos personales de
servidores en litigios laborales e información clasificada relacionada con contratos.
Asuntos administrativos bajo secreto.

El riesgo de tales fugas de información se produce a través de múltiples vectores. El primero
Es el almacenamiento en los servidores de la empresa proveedora de IA: la mayoría de las herramientas
LLM disponible en el mercado de consumo transmite las entradas del usuario a
servidores externos, donde se pueden almacenar, utilizados en el entrenamiento de versiones.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

desarrollos futuros del modelo y sujetos a fallos de seguridad. El uso de herramientas
Anuncios de IA con información confidencial sobre procedimientos sin la firma de un contrato.
Una garantía específica que asegure la confidencialidad y la eliminación de datos puede constituir una violación.
a las normas de secreto profesional y a la Ley General de Protección de Datos (LGPD, Ley No.
13.709/2018).

El segundo vector es la exposición de información durante la interacción con el modelo:
Para formular una consigna para que la modelo redactara una refutación, el fiscal
necesariamente incluye los elementos fácticos del caso, la identificación de las partes y la información.
estratégico. En sistemas sin protección de datos adecuada, esta información puede ser
al que acceden terceros: empleados de la empresa, investigadores o actores maliciosos.
en caso de una violación de seguridad.

El tercer vector es el entrenamiento del modelo basado en los datos ingresados por
usuarios. Varias herramientas comerciales de IA generativa utilizan las interacciones de
usuarios para mejorar continuamente sus modelos. Esto significa que la información
La información confidencial insertada por un abogado puede, en teoría, aparecer en las respuestas proporcionadas a...
otros usuarios, lo que constituye una grave violación de la confidencialidad profesional.

4.3 Manipulación de las respuestas de la IA

Además de las alucinaciones espontáneas y los ataques de inyección rápida, hay un tercero.
Un riesgo de naturaleza distinta: la manipulación intencional de las respuestas de la IA por parte de actores que,
Al conocer el funcionamiento de un sistema de IA en particular, estructuran su estrategia.
procedimental para influir en el análisis que realizará el sistema adversario. En contextos
en el que la parte contraria es consciente de que el Tesoro Público utiliza un determinado
Con las herramientas de IA para analizar documentos de procedimientos, surge la posibilidad, incluso más
más teórico que práctico en la etapa actual: que el partido formula sus declaraciones de una manera más teórica que práctica en esta etapa.
analizar las deficiencias del modelo utilizado por el organismo público.

Un ejemplo más inmediato y concreto de manipulación es lo que se denomina jailbreaking:
un conjunto de técnicas desarrolladas para eludir las salvaguardas éticas y de seguridad de
Los LLM, induciéndolos a producir contenido que normalmente se negarían a generar. En el contexto
En la defensa pública, un fiscal que utiliza técnicas de fuga de la cárcel para inducir al modelo.
producir argumentos que, si se evaluaran éticamente, se considerarían abusivos o

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

En contra del interés público, sería utilizar la IA como instrumento para prácticas
Prácticas procesales reprobables, con potencial para constituir mala fe procesal.

4.4 Medidas de mitigación

La adopción de medidas de mitigación eficaces requiere un enfoque por capas que
Combinar salvaguardias técnicas, procedimentales e institucionales.

En el plano técnico, las medidas prioritarias incluyen: (a) la implementación de sistemas de
Infraestructura de IA local o en una nube privada, con control total sobre los datos.
procesados, eliminando el riesgo de transmisión a servidores de terceros; (b) la selección de
herramientas que ofrecen garantías contractuales de que los datos no se utilizarán para la formación.
y la eliminación periódica de datos; c) la implementación de sistemas de filtrado de entrada
y resultados que identifican patrones de inyección rápida antes de que se publique contenido malicioso.
procesados por el modelo; (d) el uso de modelos ajustados finamente — ajustados a partir de datos
Asuntos legales gestionados por la propia agencia — para reducir el riesgo de alucinaciones en ámbitos jurídicos
aspectos legales específicos.

En términos procesales, las medidas esenciales incluyen: (a) la obligación
Verificación humana de toda la jurisprudencia y legislación generada por
(a) el modelo, mediante la consulta de fuentes primarias; (b) la prohibición de la inserción de datos.
Identificó piezas e información clasificada en herramientas que no son de IA.
aprobado por la agencia; (c) la implementación de controles de acceso que restrinjan el acceso
(d) la creación de registros de
Usos que permiten auditar posteriormente el uso de la herramienta en cada caso.

A nivel institucional, las directrices deben establecer: (a) la creación de una política
Directrices formales sobre el uso de la IA, aprobadas por el órgano rector y de obligado cumplimiento para todos los miembros;
b) la designación de una persona institucional responsable de la gobernanza de la IA (Jefe de IA)
Oficial de Cumplimiento o equivalente); (c) la provisión periódica de capacitación sobre el uso
administrador de IA; (d) la creación de un canal para informar incidentes relacionados con el uso
Insuficiencia de la IA.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

5. Ingeniería de respuesta rápida aplicada a la defensa de los intereses públicos

5.1 Concepto y relevancia de la ingeniería rápida

La ingeniería rápida es la disciplina que se ocupa de
desarrollo estratégico de instrucciones proporcionadas a modelos de lenguaje con el objetivo
para maximizar la calidad, precisión y relevancia de las respuestas generadas. Lejos de ser una
Aunque no es una actividad común, la ingeniería rápida constituye una habilidad técnica específica que
Esto puede determinar decisivamente la diferencia entre un resultado de IA genérico y uno de bajo nivel.
valor y una solución precisa, estructurada y legalmente pertinente.

En el contexto de la defensa del Estado, el dominio de las técnicas de ingeniería de
La prontitud puede transformar radicalmente la productividad del fiscal, al tiempo que...
lo que reduce —aunque no elimina— el riesgo de alucinaciones y salidas inapropiadas.
desarrollar una consigna bien estructurada que proporcione al modelo suficiente contexto,
Las restricciones claras y las instrucciones de formato específicas dan como resultado actas muy superiores a...
obtenido con indicaciones genéricas e imprecisas.

5.2 Técnicas fundamentales de ingeniería de precisión

Las principales técnicas de ingeniería de impulsos con aplicación a la práctica jurídica son:
la siguiente:

a) Indicaciones de rol: consiste en instruir al modelo para que asuma un rol.
Una indicación específica antes de comenzar la tarea. Para la defensa pública, una indicación efectiva puede comenzar
com: "Usted es un abogado estatal especializado en derecho tributario, con experiencia
en litigios tributarios y con profundo conocimiento de la jurisprudencia del Tribunal Superior de Justicia (STJ) y del Tribunal Supremo Federal (STF) sobre el
Tema. Analice el siguiente caso y elabore..." Este tipo de instrucción guía al modelo a
para acceder a patrones lingüísticos y estructurales que se aproximen más al perfil deseado.

b) Proporcionar contexto (configuración del contexto): la calidad de la salida del modelo es directamente
proporcional a la cantidad y calidad del contexto proporcionado. Una indicación eficaz debería
especificar la naturaleza del documento solicitado (defensa, apelación, opinión, memorándum), el
el tribunal receptor, la etapa procesal, los fundamentos fácticos relevantes y el
argumentos legales que el fiscal pretende respaldar.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

c) Instrucción que sugiere una cadena de pensamiento: al instruir al modelo
presentar su razonamiento de manera estructurada y secuencial antes de redactar el texto final,
El fiscal obtiene mayor transparencia respecto al camino lógico adoptado por el modelo, que
Esto facilita la identificación de errores y fallos en la argumentación. La instrucción podría ser algo así:
"Antes de redactar la defensa, presente un borrador con: (1) la identificación de las alegaciones."
(1) la demanda del demandante; (2) las excepciones procesales aplicables; (3) las excepciones sustantivas, con indicación
de los fundamentos jurídicos; (4) el orden de presentación que maximice la eficacia de
defensa."

d) Indicación negativa: así como es importante especificar qué es lo que se debe hacer
Como buen modelo a seguir, es igualmente importante enseñarles lo que no deben hacer. Ejemplos
Aplicable: "No cite jurisprudencia sin indicar el número de caso y el tribunal"; "No
Utilice argumentos que impliquen el reconocimiento implícito de los derechos de autor; No
Incluya declaraciones sobre hechos que no estén incluidos en los documentos proporcionados.

e) Instrucción de autocomprobación: el abogado puede instruir a la plantilla para que verifique
la salida real antes de presentarla. Por ejemplo: "Después de redactar la defensa, revise la
texto e identificar: (1) cualquier afirmación que no esté respaldada por los documentos
siempre que; (2) cualquier precedente citado, que indique que el abogado debe verificar su
existencia y exactitud; (3) cualquier inconsistencia lógica." Esta técnica no elimina la
alucinaciones, pero las señala, facilitando la reevaluación humana.

5.3 Estandarización institucional: bibliotecas de avisos

Una de las contribuciones más relevantes que la ingeniería de urgencia puede ofrecer a
La promoción estatal consiste en la creación de bibliotecas institucionales de indicaciones estandarizadas.
Repositorios de instrucciones cuidadosamente elaboradas, probadas y validadas para las tareas.
más frecuentes para cada órgano. La estandarización de las indicaciones ofrece múltiples ventajas:
reduce la variabilidad en la calidad de los resultados generados por diferentes proxies; garantiza
que las instrucciones incorporen las salvaguardas necesarias contra las alucinaciones; facilita la
capacitación de nuevos miembros; y permite la actualización centralizada de las instrucciones cuando
No existen cambios legislativos ni jurisprudenciales relevantes.

Una biblioteca de sugerencias para la defensa de intereses estatales podría incluir...
las siguientes categorías: (a) indicaciones para redactar objeciones en asuntos

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

(a) impuestos; (b) indicaciones para redactar apelaciones en materia de seguridad social; (c) indicaciones para la preparación de opiniones sobre la regularidad de los contratos administrativos; d) indicaciones para la investigación de jurisprudencia sobre temas específicos; (e) indicaciones para la consolidación y síntesis de argumentos en casos repetitivos; (f) indicaciones para el análisis de riesgos legales y probabilidad de éxito.

Cada pregunta estándar debe redactarse con la participación de abogados. especializado en la materia pertinente, probado en escenarios reales anónimos y sometido a un proceso de validación antes de ser incorporado a la biblioteca. Mantenimiento La gestión de la biblioteca debería ser responsabilidad de un equipo multidisciplinario que combine... Experiencia jurídica y conocimientos técnicos en inteligencia artificial.

5.4 Control de calidad de los resultados generados por IA

Independientemente de la calidad de la solicitud utilizada, toda la salida generada por la IA debe someterse a un proceso de control de calidad antes de ser incorporado a una pieza Documento procesal firmado por el abogado. Este proceso debe formalizarse en una lista de verificación. directrices institucionales que abarquen, como mínimo, los siguientes aspectos:

- Verificación de todos los precedentes citados, consultando las fuentes primarias (bases de datos). (a partir de los registros judiciales) para confirmar la existencia, la numeración y el contenido del resumen;
- Verificación de la validez de todas las disposiciones legales mencionadas;
- Evaluación de la coherencia entre los hechos narrados por el modelo y los documentos de proceso;
- Evaluación de la conformidad de los argumentos con la tesis adoptada por el órgano. institucionalmente en el asunto;
- Revisión del idioma para asegurar que se ajuste al estilo del documento legal y al lenguaje requerido. formal del cuerpo;
- Verificación de la ausencia de información confidencial que no deba incluirse en el documento;
- Evaluación de la idoneidad ética de los argumentos utilizados.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

6. USO SEGURO: APLICACIONES PRÁCTICAS EN LA DEFENSA DE INTERESES POR PARTE DEL ESTADO

6.1 Minutos de redacción

La redacción de documentos de procedimiento es la aplicación más inmediata.
y el mayor impacto de la IA en la defensa de los intereses del Estado. Para casos de naturaleza repetitiva:
Procedimientos de ejecución tributaria, impugnaciones a la ejecución de sentencias, objeciones preliminares.
aplicabilidad, impugnaciones estandarizadas en materia de seguridad social o impuestos. -, el
La plantilla puede generar, en segundos, un borrador estructurado que, cuando se envía para su revisión por
Según el fiscal, puede completarse en una fracción del tiempo que llevaría preparar todo el documento.

El protocolo recomendado para redactar actas utilizando IA debe seguir el siguiente:
los siguientes pasos: (1) el abogado identifica la categoría del documento y selecciona la solicitud predeterminada.
(2) la solicitud se personaliza con los datos correspondientes a la biblioteca institucional;
específico para el caso, respetando estrictamente la prohibición de introducir datos.
identificados en herramientas no aprobadas; (3) el borrador generado se envía a la lista de verificación
control de calidad; (4) el abogado mejora, corrige y complementa el borrador con
elementos de estrategia procesal que requieren su juicio profesional; (5) el documento
El documento final es firmado exclusivamente por el abogado responsable, sin ninguna referencia a...
La participación de la IA en su desarrollo está prohibida, a menos que la ley estipule específicamente lo contrario.

6.2 Investigación de jurisprudencia

La investigación jurídica asistida por IA presenta un potencial para mejorar la eficiencia.
particularmente significativo. Modelos equipados con acceso a bases de datos legales.
Las herramientas de búsqueda de Internet o actualizadas pueden, a partir de indicaciones
apropiado, identificar la posición de los tribunales superiores sobre un asunto determinado, mapear
la evolución de una tesis a lo largo del tiempo, identificando los juicios recientes que la han modificado.
consolidar entendimientos y sintetizar, de manera organizada, los fundamentos de
Principales precedentes sobre un tema.

Sin embargo, el riesgo de alucinaciones es particularmente alto en la investigación.
la jurisprudencia, lo que hace esencial verificar todos los precedentes.
identificado por el modelo en las fuentes primarias. El uso de la IA para la investigación de jurisprudencia.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Debe tratarse como un punto de partida, una ayuda para identificar candidatos para precedentes relevantes, y nunca como fuente definitiva de información jurisprudencial. La consolidación final siempre deberá ser el resultado de una consulta directa con las bases de datos oficiales de los tribunales.

6.3 Consolidación de argumentos

Otra aplicación de gran valor es el uso de la IA para consolidar y organizar argumentos. asuntos legales complejos. En casos de gran complejidad fáctica o normativa, el abogado Puedes utilizar la plantilla para identificar inconsistencias internas en el trabajo de la parte contraria. organizar los argumentos de la defensa jerárquicamente, identificar argumentos subsidiarios o Considere opciones alternativas que no se hayan tomado y estructure el texto de manera que se maximice la claridad y el impacto. de los principales argumentos.

Para este propósito, la técnica de dotar al modelo de, además de la aviso de instrucciones, los documentos relevantes del proceso para el análisis — señalando, más Una vez más, se impusieron restricciones a la introducción de datos confidenciales en sistemas no certificados. El análisis contextualizado de documentos mediante este modelo puede revelar elementos que... El fiscal, presionado por la carga de trabajo, podría no identificarse con lo mismo... agilidad.

6.4 Gobernanza institucional del uso de la IA

La implementación responsable de sistemas de IA en la defensa de intereses públicos requiere... la creación de una estructura de gobernanza que disciplina sistemáticamente a todos los Aspectos relevantes del uso de la tecnología. Los elementos esenciales de esta estructura son:

Política de uso de IA: un documento formal, aprobado por la dirección de la organización, que Establecer los principios que rigen el uso de la IA, las herramientas aprobadas, la categorías de información que no se pueden ingresar en los sistemas de IA, los requisitos de Verificación y supervisión humanas, y las consecuencias del incumplimiento de las directrices.

Comité de Gobernanza de la IA: un grupo multidisciplinario compuesto por fiscales, profesionales de tecnología de la información y especialistas en seguridad de la información, responsable de evaluar nuevas herramientas de IA, actualizar la política institucional e investigar reportar incidentes y recomendar mejoras al protocolo de uso.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Programa de capacitación: capacitación obligatoria para todos los fiscales sobre
Los fundamentos del funcionamiento de los másteres en Derecho (LLM) y los riesgos asociados a su uso en la práctica jurídica.
técnicas de uso público y responsable, incluyendo ingeniería y procedimientos rápidos.
verificación de salida.

Auditoría periódica: revisión regular del uso de sistemas de IA en la organización, con
evaluación de la calidad de los resultados generados, identificación de patrones de uso inapropiado y
Verificación de la efectividad de las medidas de mitigación implementadas.

7. RESPONSABILIDAD DEL ABOGADO POR EL USO DE
Contenido generado por IA

7.1 Fundamentos de la responsabilidad profesional

La cuestión de la responsabilidad del fiscal por el uso del contenido generado por
La inteligencia artificial es quizás el aspecto más delicado y menos regulado de este tema.
El punto de partida del análisis es la observación de que la responsabilidad del documento procedimental
Permanece entera y exclusivamente en poder del abogado que lo firma, independientemente de...
grado de participación de los sistemas automatizados en su desarrollo. La IA no está sujeta a
En derecho, carece de capacidad para litigar y no puede ser considerada responsable por errores.
materiales o argumentos.

La legislación brasileña no contempla, en la actualidad, ninguna disposición de este tipo.
regulación específica que rija el uso de la IA en la ley o que asigne responsabilidad por
Errores derivados del uso de sistemas automatizados. Dada esta brecha, el análisis de
La responsabilidad del fiscal debe llevarse a cabo sobre la base de principios generales que
Regulan el ejercicio de la abogacía y el desempeño de los cargos públicos.

7.2 Deber de diligencia y responsabilidad por error material

El Estatuto de la Profesión Jurídica (Ley nº 8.906/1994), en su artículo 32, establece que la
Un abogado es responsable de los actos que, en el ejercicio de su profesión, comete con dolo o negligencia.
El Código de Ética y Disciplina de la Orden de Abogados de Brasil (OAB), a su vez, consagra el deber de diligencia como...
uno de los pilares del ejercicio de la abogacía. Para los fiscales estatales, estos deberes son

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Se combinan con los principios de legalidad, imparcialidad, moralidad y eficiencia que
Rigen la Administración Pública (artículo 37, caput, de la Constitución Federal).

El uso de la IA sin la debida supervisión humana puede constituir una violación de...
Deber de diligencia en múltiples escenarios. El más obvio es la petición que cita precedentes.
inexistentes, sin que el fiscal haya verificado su existencia, una situación que, en derecho
En Estados Unidos, esto ya ha dado lugar a sanciones disciplinarias y decisiones judiciales que advierten...
Respecto a los riesgos del uso irresponsable de la IA. En Brasil, aunque aún no se han tomado decisiones...
En lo que respecta a decisiones judiciales específicas sobre el tema, los principios deontológicos predominantes son suficientes para...
para fundamentar la rendición de cuentas.

Verificación de que el fiscal utilizó inteligencia artificial para redactar el documento sin revisión.
apropiado, lo que resultó en un error material que causó daño al Tesoro Público, podría
establecer la responsabilidad en tres niveles distintos y acumulativos.

En derecho civil, la responsabilidad del abogado ante la entidad pública que
Esta representación se basa en el Artículo 37, § 6, de la Constitución Federal, que consagra la
responsabilidad objetiva del Estado frente a terceros y luego abre la posibilidad de
Derecho de recurso contra el funcionario público que causó el daño cuando dicho funcionario actuó con intención.
o culpa. La responsabilidad regresiva presupone la demostración de conducta culpable o
mala conducta deliberada por parte del fiscal: en el caso que se examina, la negligencia encarnada en la ausencia de
Revisión adecuada de la parte generada por IA —, del daño real a los fondos públicos y el vínculo causal de
causalidad entre ambos. El régimen jurídico aplicable a esta responsabilidad varía.
dependiendo de la situación laboral del fiscal: federal, estatal o funcionario público.
Los municipios están sujetos a sus respectivas leyes y regímenes jurídicos;
Los empleados públicos están sujetos al régimen CLT (Consolidación de las Leyes Laborales); y los miembros de trayectorias profesionales con estatutos específicos,
a las normas que rigen su carrera, sin perjuicio de la aplicación subsidiaria de los estatutos.
general. El denominador común de todos estos regímenes es el requisito constitucional de la intención.
o culpa como requisito previo para una reclamación de reembolso, lo que la hace legalmente relevante para los fines
En cuanto a la rendición de cuentas, la demostración de que el fiscal no ejerció la supervisión...
La responsabilidad humana que recaía sobre ellos con respecto al contenido generado por la herramienta de inteligencia artificial.

En términos disciplinarios, los fiscales no solo están sujetos a las reglas generales de
funcionarios públicos, que rinden cuentas principalmente ante los órganos disciplinarios correspondientes.
de sus carreras, según la legislación específica que las rige. Diversidad
El marco institucional de la defensa pública brasileña implica una pluralidad de regímenes disciplinarios:

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

A nivel federal, los miembros de la Oficina del Fiscal General se rigen por la ley.
Ley Complementaria N° 73/1993, que estructura la carrera, define las funciones y deberes
Establece el régimen disciplinario aplicable, con aplicación subsidiaria de prohibiciones y obligaciones.
y las sanciones previstas en el estatuto general de los funcionarios públicos federales; a nivel estatal y distrital,
Los fiscales estatales y federales están sujetos a sus respectivas leyes orgánicas.
y estatutos de defensa pública local, que por regla general asignan a las propias oficinas del defensor del pueblo
Las fiscalías tienen la facultad de investigar las infracciones cometidas por sus miembros; a nivel municipal,
Los abogados municipales están sujetos a los estatutos funcionales municipales y, cuando corresponda,
Existen reglamentos disciplinarios específicos para las Fiscalías; los fiscales de
Asambleas Legislativas, Cámaras Municipales y otros órganos del Poder Legislativo
Están vinculados a las normas disciplinarias específicas de cada institución. En todos estos casos, la edición
Una política institucional formal sobre el uso de la IA proporciona apoyo regulatorio adicional a
Iniciar procedimientos disciplinarios en caso de incumplimiento de las directrices establecidas.
haciendo que la demostración de la infracción disciplinaria sea más objetiva. Los tipos disciplinarios más comunes
Los elementos comunes a todos estos estatutos —y potencialmente aplicables al caso que se examina— son:
Negligencia en el desempeño de las funciones, incumplimiento de las normas internas.
y la práctica de actos incompatibles con la dignidad del cargo.

Es importante señalar que la responsabilidad disciplinaria presupone la demostración de
Conducta negligente o intencional. El mero uso de la IA como ayuda en la redacción de documentos.
cuestiones de procedimiento, siempre que vayan acompañadas de una supervisión y verificación humana adecuadas.
La información generada no constituye, en sí misma, ninguna infracción. La ilegalidad surge cuando...
La herramienta se utiliza de forma irresponsable, sin la debida diligencia profesional.
lo que dio lugar a un error que causó —o podría causar— daño al interés público.

7.3 Rendición de cuentas y transparencia

Una cuestión que sigue abierta tanto en la doctrina como en la práctica es la de la obligación de revelar información.
En otras palabras, si el abogado tiene el deber de informar al tribunal y a la parte contraria que ha utilizado...
Sistemas de IA en la redacción de documentos legales. Algunos tribunales estadounidenses ya...
Han emitido normas a tal efecto, que exigen a las partes declarar en una petición si han utilizado inteligencia artificial.
Generativo en el proceso de elaboración. En Brasil, el Consejo Nacional de Justicia inició el debate sobre
El tema existe, pero aún no hay ninguna norma vinculante que imponga este deber a los abogados de oficio.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Desde un punto de vista ético, existen argumentos relevantes tanto a favor como en contra.
Obligación de divulgación. A favor, se argumenta que la transparencia en cuanto al uso de la IA es esencial.
Permite que tanto el tribunal como la parte contraria comprendan el proceso de redacción del documento y ejerzan una mayor discreción.
escrutinio de su contenido. Se argumenta que la divulgación del uso de herramientas de
No se requiere ningún trabajo adicional en relación con otras herramientas tecnológicas, como los procesadores de texto.
o bases de datos legales, y que la responsabilidad del abogado sobre el documento sea completa.
independientemente de cómo se haya desarrollado.

La posición más equilibrada, en nuestra opinión, es que, en ausencia de una regla que
Si bien se impone un deber, no existe la obligación de revelar el uso de la IA como herramienta de apoyo.
Sin embargo, cuando el uso de la IA da como resultado un error identificado por el propio fiscal, la ética
El profesional exige la corrección inmediata del documento, independientemente de las consecuencias.
ya se han producido.

8. Límites éticos y legales de la delegación de actividades De los intelectuales a la IA

8.1 La insustituibilidad del juicio profesional

Un análisis de los beneficios y riesgos potenciales de la IA en la defensa de intereses estatales conduce a...
Esto conduce inevitablemente a la cuestión filosófica más profunda que subyace a todo el debate: ¿existen?
Actividades intelectuales propias de la profesión jurídica que no pueden delegarse a sistemas.
automatizado, por razones que no son meramente técnicas, vinculadas a la incapacidad actual
del modelo, pero ¿estructural, ética y legalmente?

En nuestra opinión, la respuesta es afirmativa. La profesión jurídica, como ejercicio del derecho, es...
Fundamentalmente, un ejercicio de juicio: la capacidad de evaluar, en un contexto...
Específico e irrepetible, ¿cuál es el curso de acción más apropiado a la luz de los hechos, la ley y...
valores que rigen el sistema jurídico. Este juicio no puede reducirse a un proceso.
reconocimiento de patrones, que es lo que hacen los LLM: implica una comprensión
Una interpretación genuina de las normas legales, una evaluación ética de los intereses en conflicto y una
responsabilidad por las consecuencias de la decisión.

En el contexto de la defensa del Estado, este carácter insustituible de la sentencia
El aspecto profesional es aún más marcado. El fiscal no representa únicamente los intereses del público.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

El interés inmediato de la entidad pública en una disputa: representa el interés público primario, que incluye Observar la ley, proteger los fondos públicos y respetar los derechos de los ciudadanos. Uno Ningún sistema de IA, por sofisticado que sea, no posee —ni puede poseer— eso dimensión de responsabilidad hacia el interés colectivo.

8.2 Límites éticos de la automatización

Desde un punto de vista ético, delegar actividades intelectuales a la IA plantea interrogantes. que la deontología jurídica aún está en proceso de abordar sistemáticamente.
El primer límite claro es el de la autoría intelectual: cuando firmas una obra que sabes que has creado desarrollado predominantemente por un sistema de IA sin la supervisión adecuada, el fiscal Esto atestiguaría implícitamente la autoría de una obra que no es suya, lo cual podría ser Se considera una forma de deshonestidad intelectual incompatible con la ética profesional. profesional.

Un segundo límite es el del interés del cliente; en el caso de la defensa estatal, el Interés público. El uso de la IA sin la supervisión adecuada pone en riesgo dicho interés. representados por riesgos que el fiscal podría evitar: la inclusión de argumentos Incorrectos en los hechos, la omisión de defensas relevantes y la adopción de estrategias medidas procesales inapropiadas para el caso específico. La subordinación del juicio profesional a La salida automatizada del modelo, sin una revisión crítica, viola el deber de lealtad al interés. representado.

Un tercer límite es el de la lealtad procesal. Artículo 5 del Código de Procedimiento Civil. El artículo 77 establece deberes de buena fe y lealtad procesal aplicables a las partes y a... sus abogados. La presentación de argumentos generados por IA que el fiscal no verificado —y que puede ser fácticamente incorrecto— puede constituir una violación de estos deberes, especialmente si resulta en la presentación de precedentes inexistentes o Afirmaciones sin respaldo documental.

8.3 Límites legales: qué permite y qué prohíbe la normativa vigente

Según la legislación brasileña vigente, no existe ninguna norma que prohíba expresamente... El uso de la IA en la redacción de documentos legales. Existen principios generales que...

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS - Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

interpretadas a la luz de las características específicas de la tecnología, establecen los límites de
Uso permitido.

Estos son usos claramente permitidos: el uso de la IA como herramienta de investigación y...
Se brindará apoyo para la redacción de las actas, siempre y cuando el fiscal proporcione supervisión humana.
calificados y asumen plena responsabilidad por el contenido de la pieza final; el uso de
Inteligencia artificial para el análisis y la síntesis de documentos; creación de bibliotecas de indicaciones estandarizadas.
para agilizar la elaboración de declaraciones sobre asuntos repetitivos.

Se trata de usos claramente prohibidos o éticamente cuestionables: la presentación de documentos.
Creado por IA sin revisión humana; inserción de información o datos confidenciales.
Datos personales protegidos por la LGPD (Ley General de Protección de Datos de Brasil) en herramientas de IA no aprobadas por el organismo regulador;
uso de técnicas de jailbreak para inducir al modelo a producir contenido que el
Las salvaguardias éticas del sistema lo prohibirían; y también atribuir la autoría a la IA en textos firmados.
por el fiscal sin el conocimiento del tribunal.

9 Pautas para el uso seguro por parte de los abogados

Resumiendo los elementos analizados a lo largo de este estudio, se proponen los siguientes:
Las siguientes son directrices para el uso seguro de sistemas de IA por parte de los fiscales estatales:

1. Principio de Supervisión Humana Obligatoria: todo documento de procedimiento que incluya
Cualquier participación de sistemas de IA en su desarrollo debe ser revisada minuciosamente.
por el abogado responsable antes de firmar. La revisión debe ser sustancial, no
meramente formal, e incluye la verificación de todos los elementos fácticos y legales.
y la jurisprudencia pertinente.
2. Principio de verificación de fuentes: no se permiten referencias jurisprudenciales, legislativas ni de otro tipo.
Los argumentos doctrinales generados por un sistema de IA deberían incluirse en documentos legales sin verificación.
anterior a las fuentes primarias. La presunción, a efectos de seguridad institucional, debe ser
Siempre existe la posibilidad de que la modelo haya alucinado con la referencia.
3. Principio de proporcionalidad de uso: la participación de la IA en la creación de la pieza debe
ser proporcional a la complejidad y el impacto del caso. En casos rutinarios y complejos
Dado su bajo coste, el uso de plantillas generadas por IA, con un proceso de revisión simplificado, puede resultar adecuado.
En casos complejos, con un gran impacto o con el potencial de sentar precedentes para otros.
En algunos casos, el papel de la IA debería ser más limitado y la supervisión humana más rigurosa.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

4. Principio de confidencialidad: Los abogados deben respetar estrictamente la
Directrices institucionales sobre qué información se puede introducir en los sistemas de IA.
Información clasificada, datos personales protegidos por la LGPD (Ley General de Protección de Datos de Brasil) e información sobre...
Las estrategias procedimentales delicadas nunca deben integrarse en elementos que no sean herramientas.
Aprobado por la agencia.

5. Principio de Responsabilidad Plena: el abogado debe ser plenamente consciente de que es
la única persona responsable del contenido de la pieza firmada, independientemente del grado de
La participación de la IA en su desarrollo. El uso de la IA no mitiga ni elimina la
responsabilidad profesional, disciplinaria o civil.

6. Principio de actualización continua: dada la rápida evolución de la tecnología de IA,
Los fiscales deben mantenerse al día sobre las capacidades, limitaciones y riesgos de
herramientas que utilizan, participando en programas de capacitación institucional y
Busco información actualizada sobre el tema.

7. Principio de notificación de incidentes: cualquier incidente relacionado con el uso indebido de
IA: identificación de alucinaciones en pruebas presentadas previamente, sospecha de inyección inmediata.
Las filtraciones de información deben ser comunicadas inmediatamente al comité de gobernanza.
de la agencia, para que se puedan tomar medidas correctivas.

CONCLUSIÓN

Este artículo ha demostrado que la inteligencia artificial generativa representa una
una transformación significativa e irreversible en el entorno laboral de los fiscales estatales,
con el potencial de aumentar sustancialmente la productividad de los fiscales en sus actividades.
redactar actas, realizar investigaciones de jurisprudencia y consolidar argumentos. Sin embargo,
Este potencial solo se aprovechará de forma beneficiosa si la tecnología se utiliza con
conciencia de sus riesgos, dentro de un marco institucional de gobernanza rigurosa y con
Preservación total del criterio profesional del fiscal.

Las alucinaciones de los modelos de lenguaje, el riesgo de inyección rápida, la fuga
La disponibilidad de información sensible y la posibilidad de manipular las respuestas constituyen amenazas reales.
que requieren respuestas técnicas, procedimentales e institucionales coordinadas. Ninguna
Estas amenazas por sí solas son suficientes para justificar la prohibición del uso de la IA en la profesión jurídica.
del Estado: tal postura sería tan desacertada como la adopción acrítica e irrestricta de

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

tecnología. El camino a seguir es el de una regulación responsable que maximice los beneficios y Minimizar los riesgos.

Ingeniería eficaz aplicada a la defensa de los intereses públicos y a la creación de bibliotecas. Las indicaciones institucionales estandarizadas representan instrumentos concretos para elevar el nivel de atención. calidad de los resultados generados por la IA, al tiempo que contribuye a la Estandarización de las demostraciones y protección institucional contra los riesgos identificados. Gobernanza del uso de la IA: con políticas formales, comités específicos y programas. La capacitación y las auditorías periódicas son condiciones necesarias para que la tecnología sea... incorporado de manera responsable a las operaciones diarias de las agencias estatales de asistencia jurídica.

En términos de responsabilidad, el análisis realizado nos permite concluir que la El abogado es plenamente responsable del contenido de cada documento que firma. independientemente de la participación de los sistemas de IA en su creación. La violación de los deberes de diligencia y supervisión, cuando resulten en daño al interés público, pueden para establecer responsabilidad disciplinaria, funcional y civil. El reconocimiento de esto La responsabilidad no debe inhibir el uso de la tecnología; al contrario, debe fomentarlo. Completo y técnicamente sólido.

Finalmente, cabe señalar que la regulación del uso de la IA en la defensa de intereses públicos es... Este asunto trasciende los límites de cada organismo y exige una acción coordinada a nivel ejecutivo. Federal y estatal. La emisión de directrices por el Consejo Federal de la OAB, por el Consejo El Departamento Nacional de Justicia y el propio Gobierno Federal, basándose en estudios técnicos. Análisis en profundidad y experiencia comparativa: esta es una medida urgente para Brasil. Desarrollar un marco regulatorio adecuado al desafío que plantea la inteligencia artificial. para el estado de derecho.

REFERENCIAS

BRASIL. Constitución de la República Federativa de Brasil de 1988. Brasilia, DF: Senado Federal. Disponible en https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Consultado el 1 de junio de 2026.

BRASIL. Ley N° 8.112, del 11 de diciembre de 1990. Establece el régimen jurídico de los funcionarios públicos civiles federales. Disponible en: en https://www.planalto.gov.br/ccivil_03/leis/L8112cons.htm. Consultado el 1 de junio de 2026.

BRASIL. Ley N° 8.906, del 4 de julio de 1994. Establece el Estatuto de la Profesión Jurídica y del Colegio de Abogados de Brasil (OAB). Disponible en: en https://www.planalto.gov.br/ccivil_03/leis/l8906.htm. Consultado el 1 de junio de 2026.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

BRASIL. Ley N° 13.105, del 16 de marzo de 2015. Código de Procedimiento Civil. Disponible en: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Consultado el 1 de junio de 2026.

BRASIL. Ley N° 13.709, de 14 de agosto de 2018. Ley General de Protección de Datos Personales (LGPD). Disponible en https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Consultado el: 1 de junio de 2026.

BRASIL. Consejo Nacional de Justicia. Resolución CNJ N° 332, del 21 de agosto de 2020. Este documento aborda la ética, la transparencia y la gobernanza en la producción y el uso de la Inteligencia Artificial en el Poder Judicial, y propone otras medidas. Disponible en: https://atos.cnj.jus.br/atos/detalhar/3429. Consultado el 1 de junio de 2026.

BRASIL. Consejo Nacional de Justicia. Resolución CNJ n.º 433 de 2023. Establece directrices para el uso de la inteligencia artificial por parte de los tribunales. Disponible en: https://atos.cnj.jus.br/. Consultado el 1 de junio de 2026.

BRASIL. Tribunal Superior Electoral. Acción de Investigación Judicial Electoral No. 0600814-85.2022.6.00.0000. Inspector General de Justicia Electoral: Ministro Benedito Gonçalves. Decisión del 14 de abril de 2023. Disponible en: https://www.conjur.com.br/2023-abr-18/tse-multa-advogado-peticao-baseada-conversa-chatgpt/. Consultado el 4 de junio de 2026.

BRASIL. Tribunal de Justicia de Santa Catarina. Sexta Sala Civil. Multa resultante de Jurisprudencia errónea generada por inteligencia artificial en una apelación de recuperación de bienes. Decisión del 19 de febrero de 2025. Disponible en: https://www.tjsc.jus.br/web/imprensa/-/tjsc-multa-autor-de-recurso-por-jurisprudencia-falsa-gerada-por-ia. Consultado el: 4 de junio de 2026.

BRASIL. Tribunal de Justicia de Santa Catarina. Quinta Sala Penal. Advertencia a abogado por recurso de hábeas corpus redactado por IA, con jurisprudencia inexistente. Decisión de febrero de 2025. En: MIGALHAS. TJ/SC advierte a abogado por recurso de hábeas corpus redactado por IA con jurisprudencia falsa. 2025. Ellos son Pablo, 10 de febrero Disponible en: https://www.migalhas.com.br/quentes/424313/tj-sc-adverte-advogado-por-hc-feito-por-ia-com-jurisprudencia-falsa. Consultado el: 4 de junio de 2026.

BRASIL. Tribunal Regional del Trabajo de la 7ª Región. 3ª Sala. Caso No. 0000702-38.2024.5.07.0016. Multa por litigio de mala fe y notificación a la OAB-CE (Colegio de Abogados de Ceará) sobre jurisprudencia ficticia con evidencia de generación por IA. Decisión de junio de 2025. En: MIGALHAS. TRT-7: sospecha de jurisprudencia generada por IA y multa para el abogado. São Paulo, 12 de junio de 2025. Disponible en https://www.migalhas.com.br/quentes/432475/trt-7-suspeita-de-jurisprudencia-gerada-por-ia-e-multa-advogado. Consultado el 4 de junio de 2026.

BRASIL. Tribunal Regional del Trabajo de la 12ª Región. Tribunal del Trabajo de Concórdia. Demanda laboral con petición inicial que contiene decisiones, citas doctrinales y nombre de juez inexistentes, generada por IA. Juez Daniel Carvalho Martins. Decisión de octubre de 2025. Disponible en https://portal.trt12.jus.br/noticias/autora-de-acao-e-multada-apos-advogada-inventar-jurisprudencia-e-desembargador. Consultado el 4 de junio de 2026.

BRASIL. Tribunal Regional del Trabajo de la 2ª Región. Caso No. 1001128-84.2024.5.02.0044. Multa por mala fe y notificación a la OAB-SP (Colegio de Abogados de São Paulo) por el uso de jurisprudencia ficticia generada por IA. Ponente: Juez Fernando César Teixeira França. Decisión de febrero de 2026. En: CONJUR. El uso de jurisprudencia falsa generada por IA conlleva multa y notificación a la OAB. São Paulo, 16 de febrero de 2026. Disponible en: https://www.conjur.com.br/2026-Feb-16/uso-de-jurisprudencia-criada-por-ia-gera-multa-por-ma-fe-e-oficio-a-oab/. Consultado el: 4 de junio de 2026.

BRASIL. Juzgado Regional del Trabajo de la 18ª Región. Demanda Laboral – Procedimiento Sumario N° 0001204-10.2025.5.18.0121. Multa por mala fe y carta oficial a la OAB-GO (Colegio de Abogados de Goiás) por el uso de...

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

Respecto a la jurisprudencia falsa atribuida a la IA. Decisión de mayo de 2026. En: ROTA JURÍDICA. El uso de jurisprudencia atribuida a la IA genera multa por mala fe y notificación a la OAB de Goiás. 2026. Disponible en: https://www.rotajuridica.com.br/uso-de-jurisprudencia-atribuida-a-ia-gera-multa-por-ma-fe-e-oficio-a-oab-de-goias/. Consultado el: 4 de junio de 2026.

BRASIL. Tribunal Superior de Justicia. Declaración oficial: intentos de utilizar inyección rápida. Las investigaciones se llevarán a cabo en el Tribunal Superior de Justicia (STJ). Brasilia: STJ, 20 de mayo de 2026. Disponible en: https://www.stj.jus.br/sites/portalp/paginas/comunicacao/noticias/2026/20052026-Los intentos de usar la inyección de profilaxis en la STJ serán investigados.aspx. Consultado el 4 de junio de 2026.

BRASIL. Juzgado Regional del Trabajo de la 8ª Región. 3er Juzgado del Trabajo de Parauapebas. Juicio. Expediente No. 0001062-55.2025.5.08.0130. Juez Suplente: Luiz Carlos de Araújo Santos Júnior. Decisión de 12 de mayo de 2026. En: MIGALHAS. Inyección inmediata Oculto en la petición inicial: el caso de Parauapebas. São Paulo, mayo de 2026. Disponible en: https://www.migalhas.com.br/depeso/455925/prompt-injection-oculto-em-peticao-inicial-o-caso-de-parauapebas. Consultado el 4 de junio de 2026.

BRASIL. Juzgado Regional del Trabajo de la 8ª Región. 3er Juzgado del Trabajo de Parauapebas. Juicio. Expediente No. 0001062-55.2025.5.08.0130. En: MIGALHAS. Inyección pronta en la petición inicial: el caso Parauapebas — Migalhas de Responsabilidade Civil. São Paulo, mayo de 2026. Disponible en: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/456362/prompt-injection-em-peticao-inicial-o-caso-de-parauapebas. Consultado el 4 de junio de 2026.

BRASIL. Consejo Nacional de Justicia. Resolución N° 615, de 2025. Dispone el uso de inteligencia artificial generativa en el Poder Judicial. Brasilia: CNJ, 2025. Disponible en: https://atos.cnj.jus.br/atos/detalhar/6001. Consultado el 4 de junio de 2026.

BROWN, Tom et al. Los modelos de lenguaje aprenden con pocos ejemplos. Advances in Neural Information Processing Systems, vol. 33, págs. 1877-1901, 2020. Disponible en: https://arxiv.org/abs/2005.14165. Consultado el 1 de junio de 2026.

CHUI, Michael et al. El potencial económico de la IA generativa: La próxima frontera de la productividad. McKinsey Global Institute. Disponible en 2023. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier. Consultado el 1 de junio de 2026.

CONJUR. Se investigarán los intentos de utilizar la inyección rápida en la STJ. São Paulo, 20 de mayo de 2026. Disponible en: https://www.conjur.com.br/2026-mai-20/tentativas-de-uso-de-prompt-injection-no-stj-serao-investigadas/. Consultado el 4 de junio de 2026.

PRINCIPIOS CONSTITUCIONALES DE LA IA. Seguridad y alineación de la IA: marcos clave. En: Blog de investigación de ANTHROPIC. San Francisco: Anthropic, 2023. Disponible en: https://www.anthropic.com/research. Consultado el 1 de junio de 2026.

DAFOE, Allan. Gobernanza de la IA: una agenda de investigación. Future of Humanity Institute, Universidad de Oxford, 2018. Disponible en: https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf. Consultado el 1 de junio de 2026.

DATA PRIVACY BRASIL. La pronta intervención en el Poder Judicial expone el costo de la IA sin gobernanza. São Paulo, junio de 2026. Disponible en: https://www.dataprivacybr.org/prompt-injection-no-judiciario-expoe-o-custo-da-ia-sem-governanca/. Consultado el 4 de junio de 2026.

HALLUCINATION Vectara. https://github.com/vectara/hallucination-leaderboard. Disponible en: Consultado el 1 de junio de 2026.

JI, Ziwei et al. Estudio sobre la alucinación en la generación de lenguaje natural. ACM Computing Surveys, vol. 55, n.º 12, págs. 1-38, 2023. Disponible en: https://arxiv.org/abs/2202.03629. Consultado el 1 de junio de 2026.

CLASIFICACIÓN

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google
RCMOS – Revista Científica Multidisciplinaria El Conocimiento.
ISSN: 2675-9128. São Paulo-SP.

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

JUSDOCS. Inyección inmediata en el proceso: multa, OAB e investigación STJ. 2026. Disponible en: https://jusdocs.com/blog/prompt-injection-processo-multa-oab-stj-inquerito-ia. Consultado el 4 de junio de 2026.

KASPERSKY. ¿Qué es la inyección instantánea? Cómo se puede manipular la IA y qué puedes hacer al respecto. 2026. Disponible en: https://www.kaspersky.com.br/resource-center/threats/prompt-injection. Consultado el 4 de junio de 2026.

KATZ, Daniel Martin; BOMMARITO, Michael; BLACKMAN, Josh. GPT-4 aprueba el examen de abogacía. Social Science Research Network, 2023. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233. Consultado el: 1 de junio de 2026.

MIGALHAS. La pronta inyección es una "cuestión de policía", advierte el ministro Salomão. São Paulo, junio. 2026. Disponible en: https://www.migalhas.com.br/quentes/457223/prompt-injection-e-caso-de-policia-alerta-ministro-salomao. Consultado el: 4 de junio de 2026.

MIGALHAS. Inyección pronta: STJ investiga intento de manipulación de IA en peticiones. São Paulo, mayo. 2026. Disponible en: https://www.migalhas.com.br/quentes/456424/prompt-injection-stj-apura-tentativa-de-manipulacao-de-ia-em-peticoes. Consultado el 4 de junio de 2026.

MIGALHAS. Inyección instantánea: la IA creó el vector, no la mala conducta. São Paulo, mayo de 2026. Disponible en: https://www.migalhas.com.br/amp/depeso/456010/prompt-injection-a-ia-criou-o-vetor-nao-a-ma-conduta. Consultado el 4 de junio de 2026.

MCNAUGHTON, David. Ataques de inyección instantánea contra aplicaciones integradas con LLM. arxiv, 2023. Disponible en: https://arxiv.org/abs/2302.12173. Consultado el 1 de junio de 2026.

OPENAI. Informe técnico de GPT-4. arxiv, 2023. Disponible en: https://arxiv.org/abs/2303.08774. Consultado el: 1 de junio de 2026.

Colegio de Abogados de Brasil. Código de Ética y Disciplina del Colegio de Abogados. Aprobado por Resolución N° 02/2015. Brasilia: OAB, 2015. Disponible en: https://www.oab.org.br/content/pdf/legislacaooab/codigodeetica.pdf. Consultado el 1 de junio de 2026.

Colegio de Abogados de Brasil. Consejo Federal. Recomendación N° 1, 2024. Permite el uso de inteligencia artificial generativa en la práctica jurídica. Brasilia: OAB, 2024.

OWASP — Proyecto Abierto Mundial de Seguridad de Aplicaciones. OWASP Top 10 para Aplicaciones de Modelos de Lenguaje a Gran Escala 2025. Disponible en: https://owasp.org/www-project-top-10-for-large-language-model-applications/. Consultado el 4 de junio de 2026.

PEREZ, Ethan; RIBEIRO, Marco Tulio. Ignore Previous Prompt: Attack Techniques for Language Models. arxiv, 2022. Disponible en: https://arxiv.org/abs/2211.09527. Consultado el: 1 de junio de 2026.

ROMM, Tony. Un abogado citó a ChatGPT en el tribunal. Al juez no le hizo gracia. The Washington Post, mayo. 25 de 2023. Disponible en: https://www.washingtonpost.com/technology/2023/05/27/chatgpt-ai-lawsuit-avianca-airlines/. Consultado el 1 de junio de 2026.

VASWANI, Ashish et al. La atención es todo lo que necesitas. Advances in Neural Information Processing Systems, vol. 30, 2017. Disponible en: https://arxiv.org/abs/1706.03762. Consultado el 1 de junio de 2026.

WEI, Jason et al. La estimulación de la cadena de pensamiento induce el razonamiento en modelos de lenguaje extensos. Advances in Neural Information Processing Systems, vol. 35, 2022. Disponible en: https://arxiv.org/abs/2201.11903. Consultado el 1 de junio de 2026.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.

Machine Translated by Google

Año VII, vol. 1 2026 | Envío: 04/06/2026 | Aceptado: 07/06/2026 | Publicación: 10/06/2026

WHITE, Jules et al. Un catálogo de patrones de indicaciones para mejorar la ingeniería de indicaciones con ChatGPT. arxiv, 2023. Disponible en: https://arxiv.org/abs/2302.11382. Consultado el 1 de junio de 2026.

FORO ECONÓMICO MUNDIAL. Aplicación de la IA en contextos jurídicos: oportunidades, riesgos y gobernanza. Ginebra: WEF, 2024. Disponible en: https://www.weforum.org/reports/.
Consultado el 1 de junio de 2026.

Este artículo se publica en acceso abierto bajo la licencia Creative Commons Attribution, que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.