

.UBAderecho



Directrices de uso de la IA generativa de texto y ChatGPT en la Justicia

Síntesis de principales aspectos

THOMSON REUTERS

LA LEY[™]



Dirección:

Juan Gustavo **Corvalán**, Mariana **Sánchez Caparrós** y Sebastián **Chumbita**

Equipo de investigación UBA IALAB:

Giselle **Heleg**, Carina **Papini**, Melisa **Raban**,
Antonella **Stringhini**, Marina **Petliuk Díaz**, Sol **García Gili**

Equipo de investigación Dyntec-Lab:

Marco **Rossi**, Franco **Orellana**

Diseño:

Victoria **Mafud**



Este documento es la síntesis de la “Guía de directrices para el uso de ChatGPT e IA generativa de texto en la Justicia”, en la que proponemos una serie de lineamientos y recomendaciones para un uso responsable, ético, adecuado y diligente en dicho ámbito.

El trabajo in extenso se encuentra disponible, bajo ese título, en la segunda edición del Tratado de IA y Derecho de editorial La Ley.



Cinco postulados para el uso de IA generativa en la justicia

1

Alfabetización, sensibilización y capacitación permanente.

2

Uso responsable, basado en la mitigación de riesgos asociados a los grandes modelos de lenguaje de IA generativa.

3

Human in the Loop: Control humano esencial previo y posterior. Prohibición de delegación de toma de decisiones.

4

Comprensión de las limitaciones intrínsecas y externas de los grandes modelos de lenguaje como ChatGPT (invenciones, alucinaciones, incoherencias, sesgos negativos, entre otras).

5

Evaluación de impacto estratégico según el ámbito de aplicación.

Posibles usos de IA generativa en la justicia:

Usos transversales:

Pueden darse independientemente de la etapa procesal de una causa y con relación a cualquier escrito y/o tarea judicial.

Ejemplos: búsqueda de información existente (CN, tratados, doctrina, jurisprudencia, etc.) y dentro de textos; solicitud de ideas o alternativas a problemas o conflictos judiciales planteados; síntesis de documentos; potenciación y profundización de argumentos jurídicos que se presentan como punto de partida al sistema; entre otros.

Usos específicos:

Se presentan para tareas o escritos específicos que deben proyectarse y dictarse en las distintas etapas procesales, instancias y fueros.

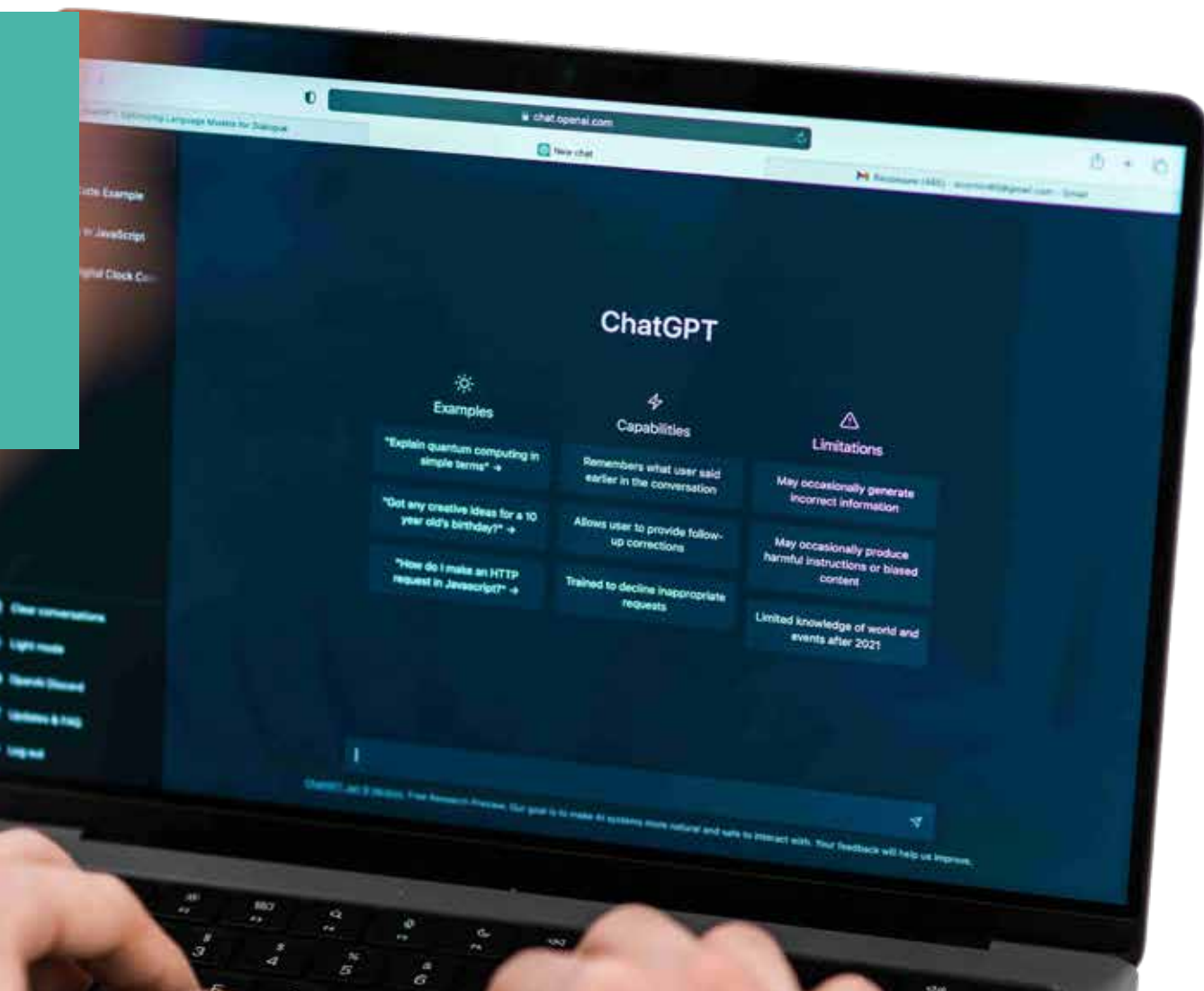
Ejemplos: redacción de borradores de providencias simples y sentencias; Relación entre textos jurídicos provistos por el usuario/a; enumeración de los medios de prueba propuestos en escritos; sugerencia de mejoras en la redacción y el contenido de textos judiciales; asistencia de redacción entre otros.

Limitaciones de la IA generativa:

Las herramientas de IA generativa arrojan respuestas coherentes y convincentes, que imitan el estilo seguro y la jerga experta. Sin embargo, tienen la capacidad de emitir respuestas **incorrectas o falsas (alucinan)**; pueden **reflejar prejuicios, estereotipos, creencias y valores sociales negativos** presentes en sus datos de entrenamiento (sesgos). **Fallan con frecuencia** cuando se les proponen tareas que implican **razonamiento lógico**; son **muy sensibles a los ajustes** en la formulación de las frases u oraciones de entrada; entre otros.

Prompting como nueva habilidad para el **coworking humano-IA:**

El ***prompt*** es una manera de comunicarse con los sistemas de IA generativa . Es importante ajustarlos y formularlos de manera adecuada para obtener mejores resultados. La habilidad para el ***prompting*** es clave para lograr resultados más eficientes, útiles, efectivos y acertados.



Síntesis de directrices y recomendaciones de implementación:

Directrices y recomendaciones	Descripción	Sugerencias
<p>Proteger y garantizar la confidencialidad de la información de la organización y de terceros</p>	<ol style="list-style-type: none">1. Implementar medidas de seguridad adecuadas para proteger la privacidad y confidencialidad de los datos de la organización y de terceros.2. Cumplir con las normas de protección de datos y las reglas de reserva del propio poder judicial.3. Tomar medidas preventivas y precautorias.	<p>Algunas sugerencias concretas:</p> <ul style="list-style-type: none">* Diseñar junto a especialistas en protección de datos y seguridad de la información, un esquema de mejores prácticas en la gestión de riesgos para garantizar la confidencialidad tanto de la información de la organización como de terceros.* Las mejores prácticas deberían incluir técnicas para que los usuarios aprendan a implementar mecanismos de anonimización y pseudoanonimización de datos para garantizar la protección de la información de la organización y de terceros.
<p>Analizar el nivel de conocimiento de los usuarios y diseñar planes de concientización y conocimiento de la IA generativa</p>	<ol style="list-style-type: none">1. La concientización y el conocimiento de las personas son fundamentales para: a) aprovechar el máximo potencial; b) garantizar un uso adecuado, ético, legal y exitoso, y c) evitar el exceso de confianza.2. Las autoridades deberían conocer cuál es el grado de conocimiento (técnico, teórico y práctico) que los agentes tienen sobre la IA generativa de texto.3. Una visión holística de su nivel de conocimiento permitirá identificar las necesidades de capacitación y abordar adecuadamente las brechas de conocimiento y de habilidad existentes.	<p>Algunas sugerencias concretas:</p> <ul style="list-style-type: none">* Diseñar e implementar campañas de comunicación específicas sobre IA generativa de texto en el Poder Judicial.* Implementar programas de capacitación y formación continua, para distintos niveles de usuarios (inicial, intermedio, avanzado) enfocados en la comprensión de la tecnología y sus implicaciones éticas.* Realizar una descripción clara de las directrices y recomendaciones para el empleo de IA generativa de texto en el Poder Judicial.* Capacitar sobre políticas y prácticas de gestión de riesgos adaptadas a la IA generativa de texto.* Enfatizar el papel de las personas en la validación y verificación de los textos generados por IA.* Establecer equipos multidisciplinarios de consulta para que puedan abordar el estudio y trabajo de casos de uso complejos.
<p>Iterar respuestas con supervisión humana adecuada enfocada en revisar y editar manualmente las respuestas</p>	<ol style="list-style-type: none">1. La supervisión humana es esencial para garantizar la precisión, calidad y ética de las respuestas generadas por IA, corregirlas y mejorarlas, prevenir sesgos, contenido inapropiado, incorrecto, inventado o ilegítimo.2. Los seres humanos seguirán siendo siempre los agentes de control y de toma de decisiones.	<p>Garantizar la supervisión humana adecuada para:</p> <ul style="list-style-type: none">* Corregir errores, mejorar la calidad y garantizar la adecuación de las respuestas.* Controlar que no contengan información falsa.* Controlar que no contengan contenido inapropiado u ofensivo.* Asegurar su adecuación ética (Ej. controlando que no reflejan estereotipos o prejuicios negativos, etc.).* Mejorar su precisión.* Aprobar metodologías para la elaboración de <i>prompts</i> adecuados a los casos.

No utilizarla como motor de búsqueda

1. Los sistemas **no están optimizados para ser empleados como un motor de búsqueda.**
2. Pueden dar lugar a la generación de **respuestas coherentes y convincentes, pero inexactas, incompletas o basadas en información desactualizada o falsa.**

Algunas sugerencias concretas:

- * Utilizar siempre mecanismos de verificación y validación humana posterior de las respuestas para garantizar su precisión y confiabilidad.
- * Utilizar siempre mecanismos de verificación y validación humana posterior de las respuestas para controlar las fuentes citadas cuando se trate de sistemas conectados a internet.
- * Tener definidas hacia el interior de la organización las responsabilidades por contenido generado que resulte incorrecto o dañino.

Considerar los documentos generados con la IA como borrador

Cuando se usa IA generativa para crear documentos judiciales, siempre se la debe utilizar **como un asistente o copiloto** generador de un borrador que agilizará la tarea de confección final del documento que, siempre, quedará a cargo de una persona humana que controlará su contenido aplicando su expertise.

Algunas sugerencias:

- * Definir lineamientos específicos para el uso de IA generativa en la redacción de documentos judiciales (Ej. en qué casos, con qué prompts).
- * Trabajar en el diseño de prompts adecuados, que le brinden al sistema un rol o perspectiva, un tono, un contexto, aporten terminología, contextualicen situaciones y brinden ejemplos, entre otros, para mejorar la calidad de los borradores generados.
- * Todos los borradores generados por IA deben ser revisados y editados por personas idóneas en la temática en la que se trabaja, para garantizar su precisión y coherencia con las normas aplicables; así como también en el formato requerido según el escrito judicial

Evaluar y mitigar sesgos en las respuestas generadas

Los sistemas de IA generativa de texto pueden **producir respuestas sesgadas por reflejar estereotipos de género, prejuicios, creencias o valores sociales negativos.**

Algunas sugerencias:

- * Capacitar y sensibilizar a los agentes del Poder Judicial sobre la problemática y sobre los distintos tipos de sesgos negativos existentes; así como de las posibles consecuencias dañinas a terceros.
- * Fomentar la colaboración entre expertos en ética de la IA y en ética judicial para desarrollar protocolos para la evaluación y mitigación de sesgos.

Directrices y recomendaciones

Descripción

Sugerencias

Ejemplos

Prompting

Un *prompt* es una **instrucción o frase inicial** que se le proporciona al modelo de lenguaje para generar texto y lograr resultados específicos en distintos formatos (Ej. noticias, correos electrónicos, poemas, etcétera).

Algunas sugerencias:

- * Ser **claro, específico** y, en la medida de lo posible, **conciso**, al proporcionar la instrucción.
- * Brindar **suficientes detalles** para que el sistema pueda procesar mejor lo que el usuario espera.
- * Fijar un **contexto adecuado** para que el sistema pueda ajustarse al tema o situación planteada.
- * Indicar **quiénes son los usuarios destinatarios**.
- * **Experimentar e iterar** el prompt.
- * Solicitar al sistema que **asuma un rol**.

1. Ej. Prompt no claro y específico:

“Escribime 15 líneas sobre IA y Derecho” / **Prompt claro y específico:** “Escribime 15 líneas sobre IA y Derecho. En particular sobre la relación que existe entre ambas disciplinas, cuáles son los puntos de contacto entre ambas y cómo se influyen recíprocamente”.

Legal prompting

Un *legal prompt* (prompt legal) es una **instrucción o frase inicial** que se le proporciona al modelo de lenguaje para la generación de **texto relacionado con asuntos jurídicos**.

Algunas sugerencias:

- *Ser **específico y claro**.
- *Establecer el **contexto**.
- *Asignar un **tono** (Ej. formal/informal).
- *Pedir un **formato** de salida.
- *Indicar un **rol**.
- *Proporcionar el **objetivo buscado**.
- *Incluir **palabras clave**.
- *Fijar **límites** (Ej. cantidad de palabras).
- *Indicar la **audiencia objetivo** (Ej. un niño o niña).
- *Solicitar la **consideración de múltiples perspectivas y contraargumentos** sobre un mismo tema.

Ejemplo de legal prompt para compliance:

Hola ChatGPT, [insertar el texto que describe los hechos del caso con suficiente detalle]. Necesito que determines la legalidad de [insertar el texto que describe los hechos a analizar con suficiente detalle], a partir de [indicar las normas bajo las cuáles se pretende efectuar el análisis].

Legal prompting para el dictado de despachos

La IA generativa de texto se puede usar como **asistente para el dictado de actos procesales** que responden a solicitudes que realizan las partes durante el proceso.

Algunas sugerencias:

- *Mantener una conversación en la que se cumplan las recomendaciones específicas para un buen prompting legal.

Prompt para generar modelos de despachos simplificados:

Hola ChatGPT. Simplifica el siguiente texto [modelo de despacho], mantén una redacción que permita que pueda ser utilizado como parte de un modelo, utiliza lenguaje jurídico, conserva datos que hagan referencia a números telefónicos, números de leyes, plazos, montos de dinero, nombres de entidades.

Super prompting

1. Un **super prompt** es un prompt que contiene varias órdenes o indicaciones juntas;
2. Permite **optimizar el diálogo** con el modelo de lenguaje;
3. Plantea la **intervención y respuesta**, en cascada y de manera conjunta, de más de un experto en una o varias temáticas.
4. Incrementa exponencialmente la productividad frente a tareas complejas como la elaboración y análisis de **mapas de procesos, mapas de tareas, árboles de decisión y matrices conceptuales**.

- * **Clarificar y segmentar** las órdenes en un mismo super prompt;
- * Respetar las **reglas de puntuación**;
- * En el primer prompt incluir **todas las órdenes** y adelantar que **la temática se va a dar en un segundo prompt**;
- * En el segundo prompt **introducir la temática**;
- * Utilizar la fórmula **“matriz conceptual”**;
- * Durante la interacción, si en modelo libera el cumplimiento de la tarea propuesta, indicar **“ok, continua”**

Hola ChatGPT. Necesito que actúes como tres expertos en ____, ____ y _____. Deberán debatir sobre el problema o cuestión que te indicaré. El debate debe ser iterado y te pido que reflejes los argumentos en un cuadro. Necesito que luego arriben a una conclusión consensuada con una propuesta o recomendación y que elabores una matriz conceptual. Luego te voy a pedir que otros tres expertos en ____, ____ y ____ pongan en crisis los argumentos y conclusión. A continuación te escribo el problema o cuestión...

Valoración de pruebas y búsqueda de recomendaciones para la resolución de casos

La IA generativa de texto puede ser utilizada:

1. Como asistente para la valoración de pruebas y hechos para obtener sugerencias o alternativas para la resolución de casos;
2. Para detectar inconsistencias o contradicciones en las pruebas y los hechos, o en los argumentos presentados;
3. Para identificar puntos clave de los hechos de un caso y extraer conclusiones;
4. Para identificar y generar contraargumentos o puntos de vista alternativos.

Algunas sugerencias:

- * Plantear un prompt informando al modelo que **solicitamos su ayuda para tomar una decisión en un caso complejo**.
- * Esbozar un **breve resumen del caso** e indicarle que **en otro prompt le informaremos las pruebas y hechos**.
- * El objetivo de los textos provistos es **obtener ayuda para la evaluación de pruebas y hechos de un caso**.

Modelo de prompt para efectuar un análisis de coherencia y consistencia entre medios probatorios. Por ejemplo, entre testimonios.

Hola ChatGPT. Te pido que revises y evalúes la coherencia lógica y argumentativa entre los testimonios A, B y C que te voy a proveer. Identifica puntos de acuerdo y puntos de contradicción entre ellos. Te pido que los vuelques en un cuadro. A continuación te doy los testimonios: [insertar el texto con los testimonios de A, B y C].

Simplificación y resumen de textos

La IA generativa de texto es útil para:

1. **resumir y simplificar textos jurídicos y no jurídicos**.
2. **identificar y extraer** información relevante, **encontrar palabras clave** y **descartar** contenido no importante.
3. **simplificar lenguaje técnico** y generar un texto expresado en **lenguaje claro** o en **lenguaje inclusivo**.

* Modelos base de prompt para simplificar un texto y proponer una salida en lenguaje claro y/o inclusivo:

1. Hola ChatGPT. Necesito que simplifiques el siguiente párrafo. Utilizá lenguaje jurídico, conservá datos que hagan referencia a números de leyes, plazos, montos de dinero, nombres de entidades o personas: [insertar el texto que quieras simplificar].