



Corte Suprema de Justicia
Secretaría de Gobierno

GUÍA DE BUENAS PRÁCTICAS PARA EL USO DE INTELIGENCIA ARTIFICIAL GENERATIVA

Esta guía establece un conjunto de directrices y recomendaciones de buenas prácticas para promover el uso responsable, ético, adecuado y diligente de la inteligencia artificial generativa en el ámbito del Poder Judicial.

La integración de la IAG en el ámbito del Poder Judicial ofrece oportunidades significativas para mejorar la eficiencia y facilitar la labor diaria, en tanto puede colaborar en mejorar la eficiencia y calidad de la ideación, análisis, redacción, comunicación y divulgación de información judicial.

Sin embargo, es fundamental que el Poder Judicial aborde los desafíos y riesgos asociados al uso de las herramientas de IAG disponibles mediante una estrategia integral que incluya: la capacitación continua de agentes, magistrados/as y funcionarios/as; la comprensión de las limitaciones intrínsecas de los grandes modelos de lenguaje; la mitigación de riesgos vinculados con esas limitaciones y el mantenimiento permanente del control humano adecuado sobre los resultados generados por IAG.

El propósito principal de esta guía es conducir al Poder Judicial en la integración de estas tecnologías, buscando un balance entre la innovación y el cumplimiento de los deberes que nos cabe, así como de principios fundamentales del derecho y la justicia. El eje rector es que la IAG constituye siempre una herramienta que complementa las habilidades humanas y no sustituye las competencias públicas.

Términos comunes.

A los fines de una mejor comprensión cabe tener presentes los siguientes términos comunes:

- **Modelo de “caja negra”**: es un término que se utiliza para referir a la falta de transparencia y explicabilidad que caracteriza muchos de los sistemas inteligentes que se utilizan en distintos campos y para diferentes fines, y que determina que sea muy complejo (y a veces imposible) explicar cómo el sistema ha arribado a un resultado específico.

- **Deep learning**: en el contexto del procesamiento del lenguaje natural, el deep learning permite al modelo aprender a entender y generar texto en lenguaje natural de manera más precisa, utilizando grandes volúmenes de datos para "aprender" las relaciones y estructuras dentro del lenguaje.

- **Inteligencia Artificial (IA)**: la definición revisada de OCDE del año 2023, describe a la inteligencia artificial como sistemas basados en máquina que “...para objetivos explícitos o implícitos, infiere(n), a partir de la entrada que recibe(n), cómo generar resultados como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad después de la implementación...”.

- **IA generativa**: es un subcampo de la IA que permite a los sistemas generar nuevo contenido en formato de texto, imágenes, sonidos, video y código informático a partir de una instrucción en lenguaje natural provista por el usuario.

- **Chatbot de IA generativa**: un programa informático que simula una conversación humana en línea utilizando IA generativa. Ejemplos



Corte Suprema de Justicia
Secretaría de Gobierno

disponibles públicamente son ChatGPT, Copilot, Gemini y Claude, entre otros.

- **Prompt**: una instrucción o conjunto de instrucciones proporcionadas en lenguaje natural a un modelo de IAG para que genere una respuesta. El prompt puede incluir información específica y directrices que guían la generación del contenido deseado.

Objetivos.

Se busca lograr los siguientes objetivos:

- Fomentar el uso responsable y ético basado en la mitigación de riesgos asociados al empleo de grandes modelos de lenguaje (alucinaciones, sesgos y discriminación, falta de transparencia y explicabilidad, riesgo de dependencia excesiva y sobreconfianza).
- Promover el control humano suficiente de las respuestas generadas por la IAG, previo y posterior, y vedar la delegación de la toma de decisiones.
- Asegurar la protección de los datos personales tratados por el Poder Judicial durante el trabajo con IAG y desplegar herramientas que contribuyan a la anonimización de documentos cuando sea necesario.
- Promover la formación y capacitación continua con perspectiva interdisciplinaria en el uso de IAG para los agentes del Poder Judicial para la comprensión de las funcionalidades, tipos de usos (ideación, análisis, redacción, corrección de contenido, entre otros), impactos potenciales, limitaciones y riesgos de los sistemas de IAG disponibles.

Directrices generales de uso.

- La implementación y uso de herramientas de IAGen, como ChatGPT, Copilot, Gemini y Claude, entre otras, no exime ni disminuye las

responsabilidades funcionales y éticas que la Constitución y las leyes establecen para magistrados/as y funcionarios/as en el cumplimiento de sus deberes. Estas herramientas tecnológicas están diseñadas para mejorar, complementar y fortalecer las habilidades y el juicio de los operadores judiciales, sin pretender sustituirlos o reemplazarlos.

- La responsabilidad de tomar decisiones judiciales recae exclusivamente en los/as magistrados/as y funcionarios/as competentes, quienes deben utilizar su criterio crítico y aplicar su conocimiento jurídico al analizar y emplear los resultados obtenidos a través de las aplicaciones de IAG.

- La IAG debe ser vista como una herramienta de apoyo que puede colaborar en mejorar la calidad del trabajo de los/as agentes, facilitar la gestión judicial, la automatización de tareas y el apoyo en procesos de toma de decisiones, sin reemplazar el análisis jurídico, la interpretación de las normas, ni la evaluación de las circunstancias específicas de cada caso, tareas fundamentales que los operadores judiciales deben continuar realizando humanamente.

- Agentes, magistrados/as y funcionarios/as están obligados a mantenerse informados sobre el funcionamiento, las capacidades y las limitaciones de las tecnologías de IAG que se implementan, con el objetivo de utilizarlas de forma ética, responsable y adecuada en el ejercicio de sus funciones.

Directrices específicas para el cumplimiento de los objetivos trazados

1. Validación obligatoria del contenido generado con IAGen: la IAG puede arrojar respuestas coherentes y convincentes que imitan el estilo seguro y la jerga experta pero que pueden ser incorrectas o falsas. En esos casos se



Corte Suprema de Justicia
Secretaría de Gobierno

suele decir que el modelo “alucina”.

Todo contenido generado por IAG debe ser revisado críticamente y contrastado con fuentes de información confiables, así como con el propio criterio y saber profesional especializado para validar que sea correcto.

2. Evaluación y mitigación de sesgos: las respuestas de la IAG pueden reflejar prejuicios, estereotipos, creencias y valores sociales negativos presentes en los datos de entrenamiento (sesgos).

Los usuarios deben poner su mayor esfuerzo para identificar y corregir contenido sesgado y/o discriminatorio en las respuestas generadas por la IAGen.

3. Transparencia y explicabilidad: las aplicaciones de IAG funcionan como “cajas negras”, lo que dificulta, cuando no impide, poder explicar el resultado que arroja el sistema frente a la instrucción dada por el usuario.

Los usuarios deben evitar el uso de IAG en ámbitos de conocimiento o para tareas sobre las que no pueden hacer una verificación con base en su propio conocimiento.

4. Dependencia excesiva y sobreconfianza: las capacidades que presentan las herramientas de IAG conllevan el riesgo de que los usuarios deleguen de manera inadvertida su capacidad crítica y de toma de decisiones. Asimismo, el estilo seguro y coherente de las respuestas dadas por la IAG puede generar una falsa sensación de seguridad en la precisión y neutralidad de las respuestas generadas, que haga que el usuario reduzca, cuando no elimine, su intervención en la verificación de los resultados.

Agentes, funcionarios/as y magistrados/as se deben capacitar sobre los riesgos asociados con el uso de IAG, para comprender sus limitaciones y

poder realizar un uso crítico y consciente, de cara a mitigar los riesgos de dependencia excesiva y sobreconfianza.

5. Mantener la seguridad: usar dispositivos de trabajo y rechazar permisos no necesarios en aplicaciones de IAG.

6. Otras limitaciones relevantes: la IAG puede fallar en tareas que requieren razonamiento lógico o que impliquen la realización de operaciones matemáticas, por lo que se debe tener especial cuidado al utilizarla en ese tipo de ámbitos.

Adicionalmente, las aplicaciones de IAG son muy sensibles a ajustes en la formulación de las instrucciones iniciales, lo que puede llevar a variaciones significativas en las respuestas. Se recomienda iterar las respuestas, no quedarse con el primer resultado obtenido y probar distintas estrategias de prompting.

Algunas aplicaciones de IAG no están conectadas a internet y poseen conocimiento limitado a cierta fecha, lo que significa que no pueden proporcionar información actualizada.

7. Supervisión humana adecuada y suficiente: los usuarios deben realizar un control humano completo y adecuado de todas las respuestas generadas por IAG, para lo cual deben tener una cabal comprensión de la temática o el ámbito de conocimiento en el que aplican la herramienta y sólo utilizarla en la medida que puedan hacer una evaluación crítica y validación de los resultados obtenidos antes de incorporarlos al proceso de toma de decisión. Se debe evitar el uso en tareas que requieren un conocimiento o juicio profesional del que se carece.

8. Prohibición de delegación: los usuarios no deben delegar la toma de



Corte Suprema de Justicia
Secretaría de Gobierno

decisiones exclusivamente en la IAG y sus respuestas no deben basarse exclusivamente en los resultados generados por la herramienta de IAG. El juicio humano debe prevalecer en todo momento, garantizando que las decisiones se tomen con un enfoque crítico y contextualizado.

9. Responsabilidad humana: agentes, funcionarios/as y magistrados/as judiciales son responsables del contenido generado con apoyo en IAG.

10. Proteger datos personales y confidenciales: el uso de IAG se debe realizar respetando la normativa vigente en materia de protección de datos personales. No se deben incluir datos personales o información confidencial en las instrucciones dadas a las herramientas de IAG cuando las políticas de privacidad aplicables no prevean que las empresas no harán uso de los datos ingresados, sea para entrenar sus modelos, como para otro tipo de usos comerciales propios de la empresa. La restricción se extiende a documentos propios de procesos de tipo no reservado.

11. Anonimización y seudonimización de datos: antes de procesar cualquier texto o documento con herramientas de IAG cuyas políticas de privacidad no prevean la protección los datos ingresados, se deben implementar técnicas y herramientas que aseguren la anonimización o seudonimización de datos personales y confidenciales existentes en dichos documentos de manera previa a utilizarlas.

12. Capacitación en protección de datos: se debe formar al personal en buenas prácticas de manejo de datos personales y confidenciales, así como materia de normativa de protección de datos de cara al uso de herramientas de IAG.

El prompting

Como se consignó, en el ámbito de la inteligencia artificial generativa de texto, un prompt o pompting es una instrucción inicial que se le da al modelo de lenguaje para que produzca texto y logre resultados específicos, como redactar textos en diversos formatos (por ejemplo, noticias, correos electrónicos, entre otros), traducir automáticamente, generar diálogos, entre muchas otras funciones.

En otras palabras, es una entrada de texto en lenguaje natural proporcionada por el usuario, que orienta al sistema en la creación de la respuesta deseada. El prompt es una forma de comunicarse con los sistemas de inteligencia artificial generativa de texto. Aunque cada modelo puede responder de manera diferente e, incluso, ante el mismo prompt, generar resultados distintos, es fundamental ajustarlos y formularlos correctamente para obtener mejores resultados.

Sugerencias específicas para el prompt:

- Ser claro, preciso y conciso: es importante proporcionar instrucciones que sean claras, específicas y, en la medida de lo posible, concisas.
- Brindar suficientes detalles: proporcionar detalles para que el sistema pueda procesar mejor la instrucción y cumplir adecuadamente con ella.
- Fijar un contexto: es útil brindar un contexto adecuado para que el sistema pueda ajustarse al tema o situación planteada.
- Indicar destinatarios: especificar quiénes son los destinatarios del texto a generar.
- Fijar un tono: indicar el tono que debe tener el texto a producir, como formal, informal, divertido, persuasivo, asertivo, etc.



Corte Suprema de Justicia
Secretaría de Gobierno

- Utilizar ejemplos y palabras clave: incluir ejemplos o palabras clave para clarificar e ilustrar el objetivo de la instrucción.
- Experimentar e iterar el prompt: en ocasiones, hacer pequeños ajustes en la instrucción o solicitar aclaraciones, precisiones y añadir nuevos detalles al resultado generado, mediante el uso de nuevos prompts, puede llevar a obtener resultados más precisos e interesantes.
- Indicar un rol: solicitar al sistema que asuma un rol específico antes de dar la instrucción.
- Solicitar un formato: especificar el formato deseado para el texto, como borrador, escrito, esquema, cuadro comparativo, diálogo, entre otros.
- Indicar un estilo de redacción: solicitar el estilo de redacción deseado para el texto a generar, como lenguaje claro, lenguaje inclusivo o inclusión de perspectiva de género.
- Cita de fuentes: para los sistemas conectados a internet (como Bing), es recomendable solicitar la cita de todas las fuentes que respalden el contenido de la respuesta generada.