

Resultados Preliminares del Programa Piloto de Uso Estratégico y Responsable de IA Generativa en la Justicia Argentina



Resultados Preliminares del Programa Piloto de Uso Estratégico y Responsable de IA Generativa en la Justicia Argentina

Objetivos

El objetivo principal de este programa piloto fue evaluar el uso responsable y estratégico de IA generativa en la Justicia para mejorar la eficiencia y calidad de las tareas judiciales. Asimismo, se buscó escalar y amplificar los resultados obtenidos en las pruebas oportunamente realizadas por el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires (UBA-IALAB) y la Junta Federal de Cortes y Superiores Tribunales de Justicia de las Provincias Argentinas y de la Ciudad Autónoma de Buenos Aires (JUFEJUS) para identificar posibilidades de aplicación de inteligencia artificial generativa (IAGen). Por último, la finalidad del programa fue también sensibilizar y alfabetizar a los integrantes de los órganos judiciales en la materia, promoviendo un uso estratégico, ético y responsable de herramientas como ChatGPT y Gemini, entre otras.

Antecedentes

A fines del año 2023, el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires (UBA-IALAB) publicó la Guía de Directrices para el uso de ChatGPT e IA generativa de texto en la Justicia con el objetivo de sentar pautas, en forma de directrices y recomendaciones, para que pudieran ser un punto de partida en relación al uso responsable de las herramientas de IA generativa de texto en la Justicia. Más recientemente, en el mes de mayo del 2024, UBA-IALAB publicó un trabajo de investigación titulado “Evaluación del impacto de la IA generativa en el trabajo”, que proporcionó un análisis detallado de la implementación de herramientas de IAGen en un amplio espectro de sectores y tareas en ámbitos clave como la Justicia y la Administración pública, entre otros.

Para el ámbito de la Justicia, UBA-IALAB y JUFEJUS realizaron un total de 29 pruebas en diferentes poderes judiciales de Argentina. Concretamente, estas pruebas fueron llevadas a cabo por equipos conformados por jueces, funcionarios y funcionarias de las provincias de Mendoza, San Juan, Tierra del Fuego, Misiones, Tucumán, Buenos Aires y la Ciudad Autónoma de Buenos Aires.

De manera global, surgió del estudio que para las tareas de complejidad alta la IAGen logró un 77% de eficiencia; para las de complejidad media, un 38% de eficiencia; y para las de baja complejidad, un 28% de eficiencia. A modo de ejemplo, una de las pruebas realizadas consistió en la identificación de los agravios de un recurso de inconstitucionalidad. Completar esta tarea sin IAGen llevaba aproximadamente 25 minutos. Con IAGen, en cambio, ese tiempo se vio acortado a 2 minutos con 30 segundos. Es decir, se logró un 90% de optimización del tiempo de trabajo.

Asimismo, de las pruebas realizadas surgió que existían desafíos en la implementación de IAGen en el ámbito judicial, pues su efectividad dependía, en gran medida, de la habilidad de los usuarios para interactuar y complementar sus tareas con estas herramientas. También encontraron limitaciones para el uso, motivadas en la necesidad de anonimizar datos para proteger la información de terceras personas involucradas en el proceso judicial, sobre todo cuando los términos de privacidad de la aplicación elegida no permitían negociar ajustes a medida.

Organizaron

La Junta Federal de Cortes y Superiores Tribunales de Justicia de las Provincias Argentinas y de la Ciudad Autónoma de Buenos Aires (JUFEJUS), el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires (UBA-IALAB), la Red Federal de Escuelas Judiciales de la Junta Federal de Cortes y Superiores Tribunales de Justicia de las Provincias Argentinas y de la Ciudad Autónoma de Buenos Aires (ReFleJar), el Instituto Federal de Innovación, Tecnología y Justicia de la Junta Federal de Cortes y Superiores Tribunales de Justicia de las Provincias Argentinas y de la Ciudad Autónoma de Buenos Aires (IFITEJ), el Laboratorio de Innovación Judicial del Poder Judicial de la Provincia de Mendoza (JusMendoza.Lab), el Laboratorio de Innovación Tecnológica de la Facultad de Derecho de la Universidad Nacional de Lomas de Zamora (UNLZ-LAB), el Laboratorio de IA, Innovación y Transformación Digital de la Facultad de Derecho de la Universidad Nacional de Tucumán (DYNTEC), el Laboratorio de Investigación y Desarrollo de Software e Inteligencia Artificial de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba (LIDeSIA), la Cátedra de Inteligencia Artificial, Tecnología y Derecho de la Facultad de Derecho de la Universidad Nacional de Mar del Plata (UNMDP), el Departamento de Derecho de la Universidad Nacional del Sur (UNS), la Facultad de Derecho de la Universidad de Palermo (UP), la Facultad de Derecho y Ciencias Sociales y Políticas de la Universidad Nacional del Nordeste (UNNE), la Universidad de Champagnat (UCH) y el Departamento de Derecho Judicial de la Universidad Austral.

Colaboraron

El Ministerio de Justicia de la Nación, a través del Programa Nacional Integral de Inteligencia Artificial en la Justicia, el Colegio Público de la Abogacía de la Capital Federal (CPACF), la Red de Innovación Juslab, la Asociación de Impulsores de Legaltech Argentina (ALTA), la Asociación de Mujeres Jueces de Argentina (AMJA), el Laboratorio de Inteligencia Artificial e Innovación del Observatorio de Derecho, Tecnología e Innovación (ODTI) de la Universidad del Museo Social Argentino (UMSA), Thomson Reuters - La Ley, la Red Mujeres para la Justicia, el Centro de Estudios en Tecnología y Sociedad de la Universidad de San Andrés (CETyS), Justicia Digital Porteña, Abogados Digitales 4.0, la Asociación de Magistrados y Funcionarios de la Justicia de Neuquén (AMyF), el Centro de Funcionarios Judiciales de Tucumán (CFJ), el Colegio de Abogados de Rosario (COLABRO), la Universidad Nacional del Comahue (UNComa), E-Procesal, Sherlock Legal y El Dial.com, la Federación Argentina de la Magistratura y la Función Judicial (FAM), la Asociación Argentina de Derecho Procesal (AADP), el Programa IAG, Derecho y Justicia de la Universidad Nacional de Tucumán, Iguales IA & Género, la Asociación de Magistrados y Funcionarios de la Justicia Nacional (AMFJN), la Procuración General de la Ciudad Autónoma de Buenos Aires, el Observatorio Federal de Seguridad y Justicia (O.FE.S.J.), el Centro de Estudios en Derecho, Innovación y Tecnología de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata (CEDlyT), el Ministerio de Justicia y Derechos Humanos de Santiago del Estero, el Consejo de Procuradores, Fiscales, Defensores y Asesores Generales de la República Argentina, la Oficina de Innovación del Tribunal Superior de Justicia de la Ciudad Autónoma de Buenos Aires, ODR Online Dispute Resolution Latinoamérica, la Universidad de San Pablo de Tucumán, la Asociación de Magistrados de Tucumán, el Colegio de Magistrados y Funcionarios del Poder Judicial de la Provincia de Buenos Aires, el Instituto de Formación e Investigación de la Magistratura, la Función Judicial del CMFBSAS y la Asociación Argentina de Justicia Constitucional (AAJC), el Consejo de la Magistratura del Poder Judicial de la Ciudad Autónoma de la Ciudad de Buenos Aires y la Revista DPI Cuántico - Derecho para Innovar.

Referentes regionales

Para asegurar una implementación efectiva y coordinada del programa, se designaron referentes regionales, los cuales fueron titulares de organismos judiciales en las principales regiones de Argentina: Noroeste, Nordeste, Centro, Cuyo y Patagonia. Estas personas referentes desempeñaron un papel clave en la orientación y la implementación del programa.



Duración del programa

Etapa 1

Bases metodológicas. Alfabetización y sensibilización inicial.

En esta etapa, se establecieron las bases metodológicas, incluyendo la metodología de medición, la selección de tareas y subtareas, así como los correspondientes métodos para analizar los resultados. Además, se abordaron otras cuestiones relacionadas con el tiempo de revisión, las cuestiones vinculadas a la protección de datos, los estándares y las buenas prácticas de uso. Aquí se incluyó una breve descripción de las directrices y una propuesta de estándares éticos y de buenas prácticas en materia de uso de IAGen.

Etapa 2

Desarrollo de las pruebas y recolección de interacciones.

En la segunda fase, las personas que participaron de la prueba piloto desarrollaron pruebas de uso de IAGen en tareas relacionadas con la función judicial, siguiendo la metodología propuesta por UBA-IALAB y JUFEJUS, con el objetivo de escalar y amplificar los resultados previamente recabados en la investigación sobre el impacto de la IA generativa en el trabajo referido en los antecedentes. Quienes participaron aplicaron técnicas de prompting y estrategias de mitigación para obtener resultados eficientes y relevantes de las herramientas de IAGen en tareas jurisdiccionales concretas.

Etapa 3

Sistematización de resultados.

Una vez completado el período de las pruebas piloto, se procedió a la sistematización de los resultados obtenidos. Esta fase implicó la recopilación de todos los datos generados. Asimismo, se analizaron lecciones aprendidas y se detallaron desafíos y potenciales riesgos asociados.

Etapa 4

Documentación y Difusión.

Finalmente, el programa avanzó sobre la documentación y difusión de los resultados, con el objetivo de que el conocimiento adquirido se comparta ampliamente. Para ello, se realizó un informe preliminar dirigido a fomentar la adopción de IAGen en la Justicia de manera estratégica, ética y responsable, que permita, a su vez, avanzar sobre otros contextos judiciales que no formen parte del programa.

Etapa 5

Confección de Directrices y Modelo de Acordada Reglamentaria

En la próxima fase del programa, se avanzará en la confección de una Guía de Recomendaciones y Directrices para el uso adecuado, ético, estratégico y responsable de la Inteligencia Artificial Generativa y de agentes artificiales en la Administración de Justicia. Este proceso buscará consolidar los aprendizajes obtenidos en las etapas previas para formular lineamientos claros que orienten su implementación en los distintos Poderes Judiciales del país. Asimismo, se trabajará en la elaboración de un modelo de acordada reglamentaria que sirva como referencia para la adopción institucional.



Resultados del programa

Resultados

Durante las pruebas piloto realizadas en distintas jurisdicciones, fueros e instancias, se recopilaron datos sobre el desempeño de herramientas como ChatGPT y Gemini. Para ello, se proporcionaron a las personas voluntarias dos archivos Excel en los que debían registrar sus interacciones con las herramientas de IA generativa, incluyendo las consultas realizadas, las respuestas obtenidas y cualquier otra información relevante. El objetivo de esta recolección de datos fue evaluar la utilidad de las herramientas en diferentes tareas judiciales, identificar sus limitaciones y analizar su impacto en la eficiencia y la calidad del trabajo judicial.

En total, se registraron 2845 interacciones en base a 1007 planillas, lo cual ha permitido observar el desempeño de los agentes conversacionales más conocidos en una variedad de organismos de todo el país, desde juzgados de Primera Instancia hasta Superiores Tribunales. Dentro de este marco, se evaluó su uso en distintos fueros, tales como el Civil y Comercial, Familia y Minoridad, Penal, Contencioso, Tributario y Consumidor, así como en ámbitos de ejecución y en organismos con competencia múltiple. La diversidad de instancias y fueros analizados permitió obtener una visión amplia sobre las áreas en las que la IAGen está siendo más empleada y aquellas en las que presenta mayores desafíos.

Asimismo, el estudio ha contado con la participación diversa de agentes judiciales, permitiendo evaluar cómo estas herramientas son utilizadas por personas en distintos roles. Se han incluido en el análisis funcionarios, funcionarias, secretarios, secretarias, jueces, juezas y personal administrativo, cuyas experiencias han sido fundamentales para evidenciar el impacto real de estas herramientas en la práctica cotidiana.

En cuanto a los agentes conversacionales utilizados, se han analizado los modelos principales de inteligencia artificial generativa vigentes al momento de las etapas de recolección de datos. ChatGPT ha sido evaluado en sus versiones 3.5, 4 y 4o, mientras que Gemini ha sido probado en sus versiones Gemini 1 y Gemini Ultra. En el caso de Copilot, quienes participaron no especificaron versiones, por lo que se presume el uso de la versión estándar integrada en los productos de Microsoft. Esta variedad de modelos y versiones, a su vez, ha permitido analizar cómo la evolución de estas tecnologías impacta en su precisión, eficiencia y capacidad de adaptación a las necesidades del ámbito judicial.

Las tareas realizadas durante las pruebas han sido diversas y han abarcado múltiples aspectos de la práctica tribunalicia. Entre las más frecuentes se encuentran la redacción de documentos, como resoluciones, dictámenes y sentencias, la extracción de datos, el análisis y búsqueda de jurisprudencia, la identificación de casos relevantes; la elaboración de resúmenes de fallos, normas y doctrina; la generación de respuestas a preguntas jurídicas específicas, evaluando la fundamentación legal de las respuestas; así como también la planificación y organización del trabajo, incluyendo la distribución de tareas y la gestión del tiempo.



Para medir el desempeño de las plataformas, se solicitó a los participantes que otorgaran valoraciones en dos dimensiones clave: la calidad del resultado, es decir, si la respuesta generada por la IAGen era precisa, clara y útil para la tarea solicitada, y la eficiencia en términos de tiempo, evaluando cuánto se redujo el tiempo de trabajo gracias al uso de la herramienta. Las puntuaciones fueron asignadas en una escala del 1 al 5, donde 1 representaba una valoración muy baja, indicando que el resultado fue poco útil o erróneo, y 5 representaba una valoración muy alta, señalando que la respuesta fue precisa y efectiva.

La valoración fue altamente positiva, promediando en todas las tareas valores mayores a 4 puntos, otorgando como calificación:

“Rápido, llevó menos tiempo de lo esperado”

Los primeros análisis indican que ChatGPT-4o y Gemini Ultra han obtenido las mejores puntuaciones tanto en calidad como en eficiencia en tiempo, mientras que versiones anteriores de estos modelos y Copilot han mostrado un desempeño más variable. Este hallazgo sugiere que las mejoras en los modelos de IA han tenido un impacto positivo en su capacidad para generar respuestas más precisas y en su utilidad en tareas específicas dentro del ámbito judicial.

	Tiempo	Calidad
Chat GPT	4,30	3,78
Copilot	3,81	3,63
Gemini	3,97	3,33
Otros	4.02	3,79

Valoraciones según cada herramienta (del 1 muy bajo al 5 muy alta)

Además de las puntuaciones numéricas, se pidió a las personas voluntarias que describan cualitativamente si notaban una diferencia sustancial en el tiempo de trabajo con y sin IAGen. En la mayoría de los casos, suscribieron que el uso de las herramientas permitió acelerar la ejecución de tareas, especialmente en la redacción de documentos y el análisis de jurisprudencia. No obstante, algunas otras señalaron que el beneficio en términos de tiempo depende del tipo de tarea y de la necesidad de verificar la precisión de las respuestas generadas. Al mismo tiempo, marcaron que en ciertos escenarios, como el análisis de documentos jurídicos complejos o la generación de respuestas a cuestiones altamente especializadas, los sistemas inteligentes demostraron dificultades, requiriendo correcciones o validaciones adicionales.

El análisis de los comentarios permitió identificar tendencias en la percepción del uso de estas herramientas. Entre los aspectos positivos más destacados se encuentra la rapidez y eficiencia con la que las herramientas pueden completar tareas repetitivas, la utilidad en la redacción de documentos y la facilidad de uso, que permite que sean utilizadas sin necesidad de una capacitación extensa. Sin embargo, también han surgido preocupaciones respecto a la necesidad de supervisión, ya que las respuestas generadas debían ser verificadas antes de ser utilizadas en el proceso judicial, la presencia de errores en la interpretación de documentos jurídicos y las limitaciones en respuestas especializadas, donde todavía puede generar información imprecisa (alucinaciones).

Se destacó la necesidad de supervisión. Las personas voluntarias resaltaron la importancia de verificar las respuestas antes de su uso, ya que identificaron errores en la interpretación de documentos jurídicos y limitaciones en la generación de respuestas especializadas.

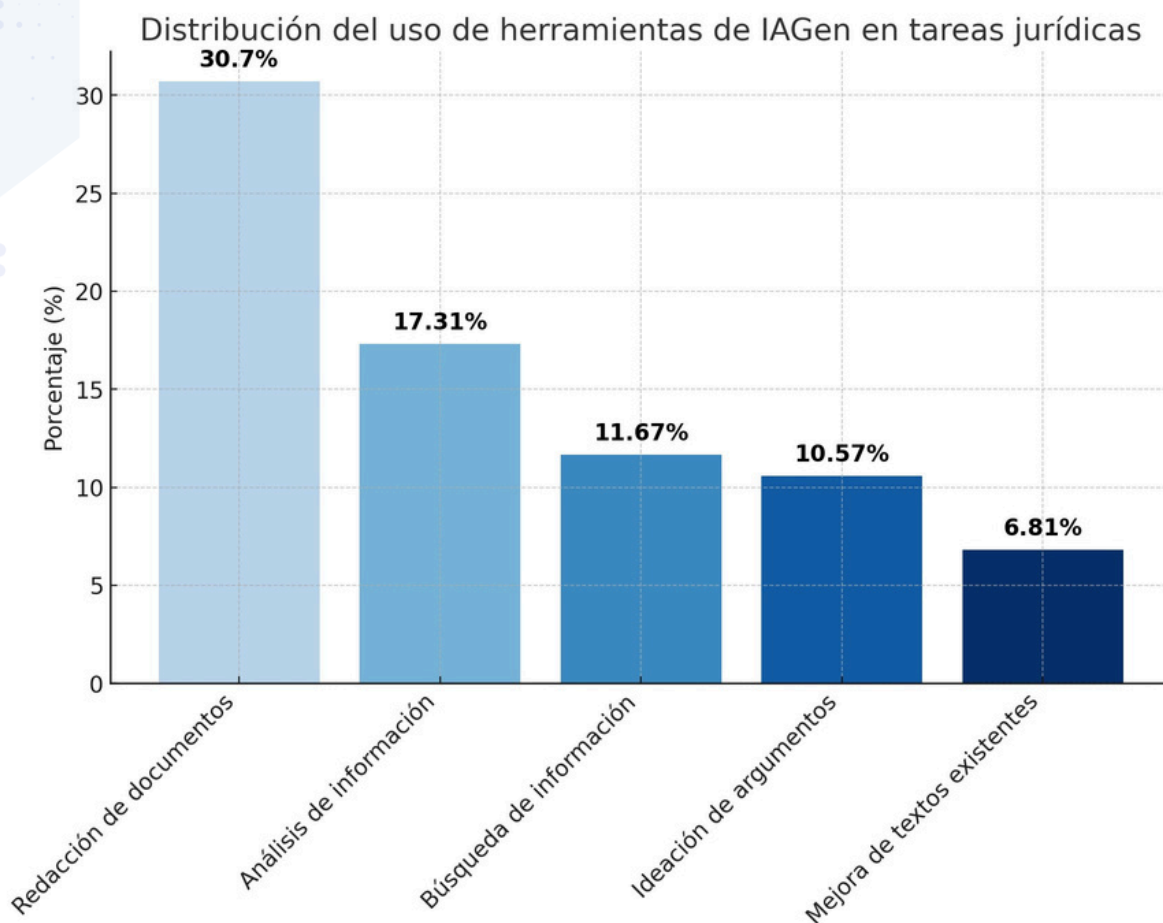
Estos datos permiten concluir que, si bien la adopción de estas herramientas se encuentra en una etapa de prueba piloto, la mayoría de los y las participantes han reconocido beneficios significativos en términos de rapidez y productividad. No obstante, la fiabilidad de las respuestas generadas y la necesidad de supervisión continúan siendo factores críticos a tener en cuenta.

Las herramientas como ChatGPT o Gemini, si bien ofrecen un apoyo considerable, no reemplazan la actividad humana, mucho menos sin revisión, ya que pueden generar graves errores que, si no son detectados, podrían comprometer la calidad del servicio de justicia y afectar los derechos de las personas.

Continuando con el análisis, cabe destacar que se trazaron métricas concretas, comparando detalladamente las valoraciones de cada modelo en función de su calidad y eficiencia. También se analizaron patrones en la adopción de la inteligencia artificial según provincia, fuero e instancia, permitiendo identificar en qué ámbitos su impacto es más significativo. Asimismo, se realizó un análisis cualitativo más preciso de los comentarios de las personas voluntarias, con el objetivo de detectar preocupaciones recurrentes y oportunidades de mejora en la implementación de estas herramientas.

Del análisis global de las interacciones, se desprende que las tareas más utilizadas con IAGen en la justicia fueron la redacción de documentos (**30,70%**), seguida del análisis de información (**17,31%**) y la búsqueda de información (**11,67%**). La ideación de argumentos y estructuras jurídicas representó el **10,57%**, mientras que la mejora de textos existentes alcanzó el **6,81%**.

Si bien este análisis se corresponde con las instancias de la pirámide judicial, se observa que en cada universo en particular, como ser el de Primera Instancia, la redacción fue la tarea predominante con 32,80%, seguida del análisis con **15,87%**, la búsqueda con **11,22%** y la ideación con **10,11%**.



También se pudo observar que se utilizó para mejoras de textos **7,19%** y profundización de argumentos/ideas **6,39%**, con un menor uso en combinación de información/documentos **4,77%**, interpretación **4,53%** y valoración **2,60%**.

En Segunda Instancia, el análisis fue la tarea más frecuente con **26,26%**, seguida de redacción con **24,82%** y búsqueda con **11,15%**. La ideación tuvo un **8,27%**, mientras que las mejoras a textos y la profundización de argumentos/ ideas alcanzaron un **6,83%** cada una. La interpretación llegó al **5,76%**, con un menor uso de combinación de información/documentos **4,32%** y valoración **3,24%**.

En Casación, la redacción fue la tarea más utilizada, representando el **33,33%** del uso, seguida por el análisis con **30,00%** y la búsqueda de información con **13,33%**. La combinación de información/documentos y la ideación tuvieron un uso menor, con **6,67%** cada una.

En los Superiores Tribunales, la redacción lideró con **25,27%**, seguida del análisis con **19,23%** y la ideación con **11,54%**. La búsqueda de información y la profundización de argumentos representaron **8,24%** cada una, mientras que las mejoras a textos fueron el **7,69%** del uso. La interpretación alcanzó el **6,04%**, y tareas como combinación de información/documentos **3,85%**, distinciones conceptuales **3,30%** y valoración **2,75%** tuvieron un uso menor.

Finalmente, en aquellos organismos identificados como de Única Instancia, la redacción representó el **66,67%** del uso de IAGen, mientras que el análisis y la profundización de argumentos tuvieron un **16,67%** cada una.

En el análisis de las distintas etapas procesales, se observa que en las etapas iniciales, la tarea más utilizada fue Redacción (**41.41%**), seguida de Comparación (**25.00%**) y Análisis (**9.49%**).

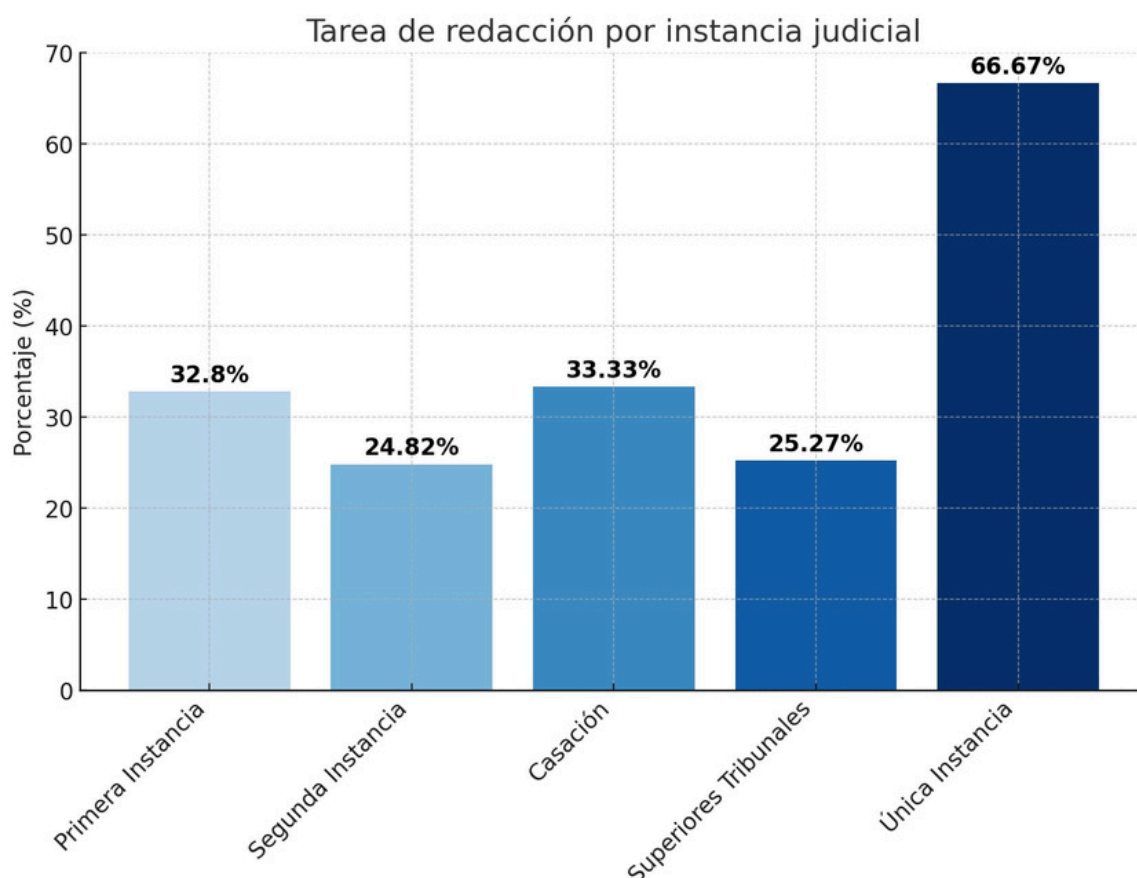
Durante las etapas probatorias y de análisis, la actividad predominante fue Redacción (**38.62%**), junto con Análisis (**14.31%**) y Búsqueda (**12.07%**), reflejando el trabajo sobre la valoración de las pruebas.

En las etapas decisorias, la IA se utilizó principalmente para Redacción (**31.54%**), con un uso también destacado en Análisis (**12.96%**) y Profundización de Argumentos/Ideas (**10.85%**), lo que indica su incidencia en la elaboración de resoluciones y sentencias.

En las etapas impugnativas, la tarea más realizada fue redacción de sentencia (**43,19%**) y Análisis (**17.05%**), evidenciando el uso de la IAGen en la revisión de agravios.

Para la etapa de ejecución, la IA se destinó en mayor medida a Redacción (**38.14%**), con una participación importante en Análisis (**21.19%**) y Combinación de Información/ Documentos (**7.63%**), facilitando la implementación de lo resuelto.

Finalmente, en la categoría Otras, la tarea más frecuente fue Redacción (**27.55%**), junto con Ideación (**16.53%**) y Análisis (**14.92%**).



Agudizando el estudio, se hizo hincapié en el análisis del contenido de los prompts que fueron compartidos. Para efectuarlo, se consideró realizar una clasificación temática basada en frecuencia y contexto, es decir, un enfoque que combinó análisis de contenido cuantitativo y agrupación semántica a partir de una “bolsa de palabras” propias del lenguaje jurídico-judicial.

Los agrupamientos fueron de acuerdo a las siguientes categorías:

A

Interpretación y aplicación del derecho (interpretación de normativas y leyes, análisis comparativo de normativas y la evaluación de principios generales del derecho)

B

Redacción de documentos (redacción de sentencias, resoluciones y dictámenes)

C

Valoración probatoria (análisis de prueba en general, interpretación de declaraciones testimoniales y redacción de fundamentos probatorios en sentencias)

D

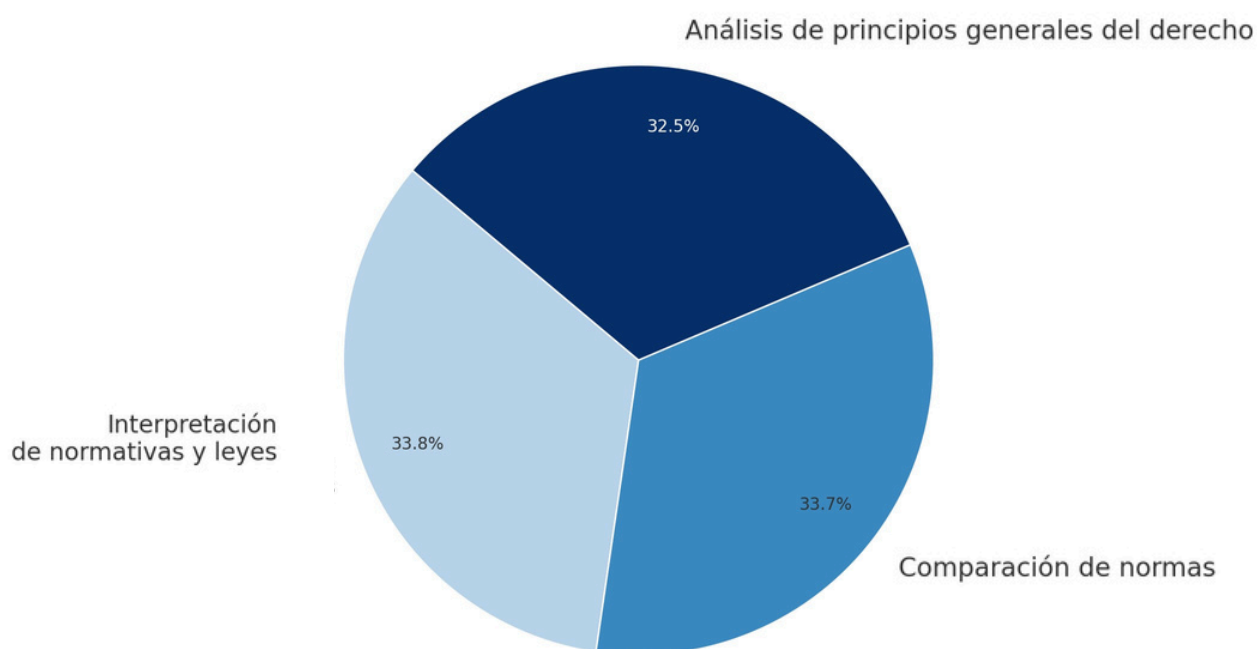
Materias de derecho consultadas

E

Búsqueda y análisis de información (jurisprudencia en casos análogos, revisión de doctrina, normas y la evaluación de fallos presentados en recursos)

Bajo tales parámetros, se pudo observar que en interpretación y aplicación del derecho, el **33.8%** de los prompts se centró en la interpretación de normativas y leyes, mientras que el **33.7%** comparó normativas entre sí para analizar diferencias legales, y el **32.5%** abordó principios generales como la buena fe y el abuso del derecho.

Distribución de prompts por categorías



En la redacción de documentos jurídicos, se enfocó principalmente en la redacción de sentencias (**33.3%**), seguida por la elaboración de resoluciones sobre medidas cautelares (**33.3%**) y la redacción de dictámenes (**33.3%**).

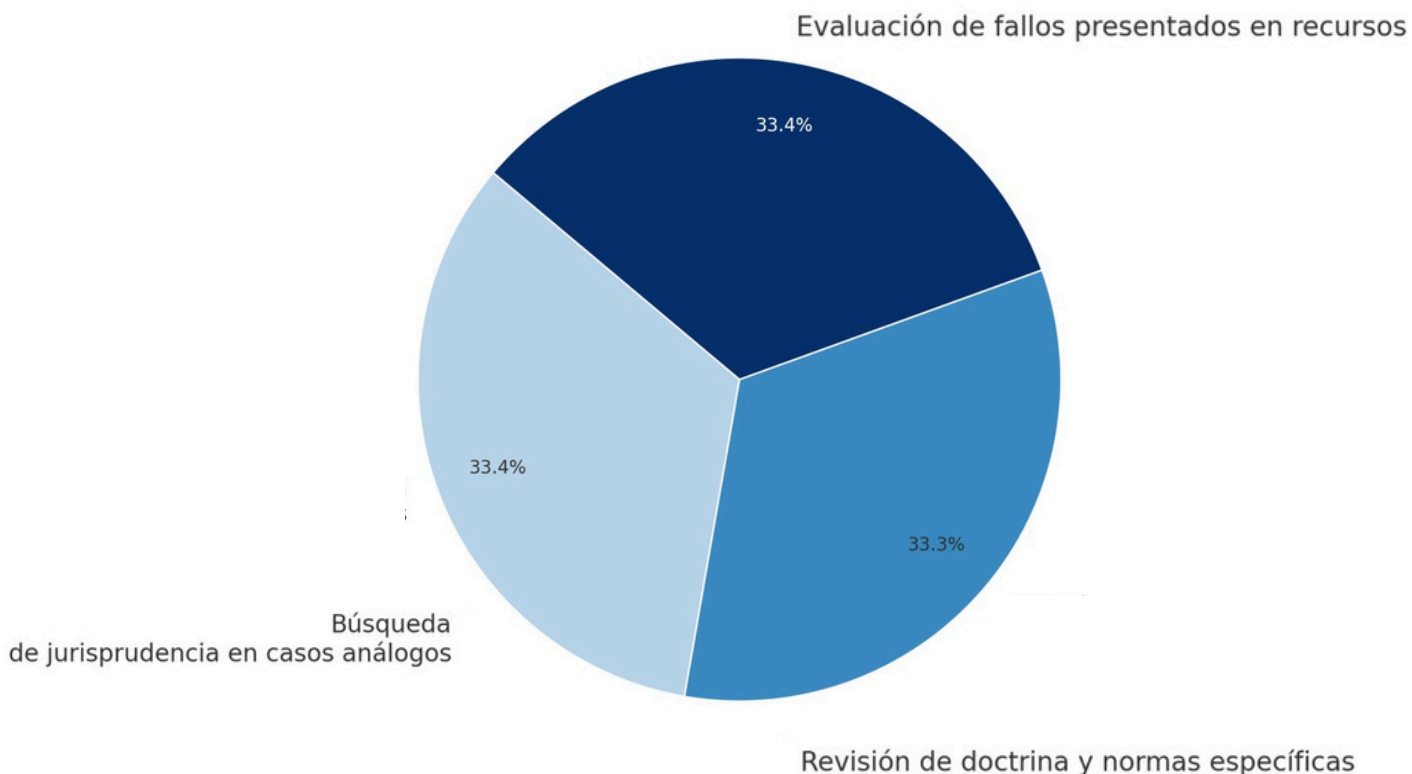
En cuanto a la evaluación probatoria, incluyó en un **33.4%** el análisis de pruebas en general, en un **33.2%** la interpretación de declaraciones testimoniales, y en un **33.4%** la redacción de fundamentos probatorios en sentencias.

Sobre las materias más consultadas, cabe decir que fueron la civil (**41.5%**), penal (**24.6%**) y comercial (**33.9%**), particularmente en responsabilidad, incumplimientos contractuales y delitos penales.

Y finalmente, en referencia a la búsqueda y análisis de información, se distribuyó en un **33.4%** para la búsqueda de jurisprudencia en casos análogos, un **33.3%** en la revisión de doctrina y normas específicas, y un **33.4%** en la evaluación de fallos presentados en recursos.

Esto sugiere que la inteligencia artificial generativa no está siendo utilizada solo como una herramienta de consulta, sino como un apoyo integral en el proceso de redacción sustancial. Su implementación permite optimizar no sólo la recuperación de información, sino también su interpretación y aplicación en contextos específicos, lo que evidencia su potencial para transformar la práctica jurídica al facilitar un acceso más eficiente y estructurado a las fuentes del derecho.

Asistencia en búsqueda y análisis de información





Conclusiones y perspectivas futuras



Conclusiones y perspectivas futuras

El Programa Piloto de Uso Estratégico y Responsable de IA Generativa en la Justicia Argentina ha evidenciado el potencial de la IAGen para transformar la práctica judicial. Los resultados obtenidos demuestran que esta tecnología puede optimizar tiempos, mejorar la calidad de los documentos y permitir a los y las agentes judiciales concentrarse en tareas más complejas que requieren su experiencia y criterio.

Este programa no solo ha implicado una investigación y análisis riguroso, sino que también es un ejemplo de colaboración entre instituciones académicas, judiciales y tecnológicas. Dicha sinergia es fundamental para asegurar que la IAGen se integre en el Poder Judicial de manera adecuada. La creación de espacios de diálogo y trabajo conjunto entre los diferentes actores involucrados permitirá afrontar los desafíos y aprovechar al máximo las oportunidades que ofrecen los agentes conversacionales.

Para eso, será necesario **continuar desarrollando estrategias** de implementación que garanticen su uso responsable y ético. Aquí la formación continua de los operadores y las operadoras judiciales en el uso de estas herramientas, así como la adaptación de las normas y procedimientos a esta nueva realidad, serán claves para asegurar una nueva forma de administrar justicia.

En línea con lo anterior, es importante destacar que durante el período de la prueba piloto, el uso de la IA generativa en el ámbito judicial se caracterizó por transitar un paradigma de trabajo en el que el acceso a esta tecnología se encontraba disponible, casi exclusivamente, a través de plataformas especializadas.

Cabe recordar que herramientas como ChatGPT, Copilot o Gemini requieren que los usuarios ingresen de forma activa a entornos específicos ajenos a las aplicaciones tradicionalmente utilizadas en el Poder Judicial. Esta realidad evidenció la clara separación entre los sistemas de gestión oficial y las soluciones basadas en IA generativa.

A ello se suma que, en la mayoría de los casos, los agentes y funcionarios acceden a estas aplicaciones bajo esquemas de licencia gratuitos o contratados de manera personal, sin intermediación de los Poderes Judiciales.

Sin embargo, particularmente a partir del último trimestre de 2024, este esquema experimentó una transformación significativa. La integración progresiva de la IA generativa en aplicaciones de uso cotidiano del Poder Judicial —como procesadores de texto (Microsoft Word, Google Docs), hojas de cálculo (Excel, Google Sheets) y plataformas de reuniones virtuales (Zoom, Teams)—, con funcionalidades impulsadas por IA que muchas veces se activan de forma nativa y automática, sin que el usuario tenga que intervenir explícitamente, ha ocasionado un cambio en aquel paradigma. Ahora el usuario tiene disponible la IAGen y puede utilizarla a través de herramientas que ya conoce y emplea para su labor diaria.

Un ejemplo paradigmático es la incorporación de Copilot en Microsoft Word, el cual ofrece sugerencias y mejoras en la redacción en tiempo real, transformando el proceso tradicional de elaboración de textos y contribuyendo a una mayor eficiencia sin requerir una activación manual.

Este avance resalta la convergencia entre la tecnología y las herramientas habituales, rompiendo la barrera que separaba la IA generativa del entorno de trabajo cotidiano del Poder Judicial.

Ambos paradigmas (el del acceso a la IAGen mediante aplicaciones específicas y el de su aprovechamiento a través de herramientas de ofimática cotidiana) conviven en la actualidad. No obstante, todavía existe un universo de agentes y funcionarios que no logran incorporar la IA para desarrollar su trabajo porque no saben cómo aprovecharla del todo o porque temen cometer errores.

Frente a esta situación, resulta imperativo repensar la regulación de la IA generativa en el ámbito judicial. Las disposiciones que promueven la necesidad de autorizaciones específicas o la obligación de informar de manera explícita sobre su uso, podrían volverse obsoletas en un contexto en el que la tecnología está integrada por defecto en las herramientas de trabajo.

Por tanto, este cambio de paradigma puede verse como una oportunidad para desarrollar estrategias regulatorias basadas en criterios de razonabilidad y proporcionalidad, que consideren el nuevo escenario y aborden el uso responsable y las medidas de prevención y mitigación de riesgos considerando variables específicas, propias de cada caso, como el tipo de acto procesal, la relevancia de los derechos implicados y el nivel de estandarización de las tareas. De este modo, por ejemplo, la normativa podría partir por establecer distinciones entre los escenarios en los que la IA generativa puede tener importante incidencia en los derechos por las características del proceso y aquellos en los que desempeña un rol más auxiliar.

En el primer caso, como podría suceder con la redacción de sentencias en causas penales, donde la IA podría sugerir argumentaciones basadas en precedentes, pero su aplicación exige una intervención humana decisiva para interpretar el contexto probatorio y garantizar el debido proceso, el impacto en derechos fundamentales requiere salvaguardas robustas frente a posibles sesgos algorítmicos o errores de interpretación.

En contraste, frente a procesos más estandarizados, como puede suceder con los juicios monitorios, la IA puede optimizar la eficiencia en la gestión documental o la generación de resoluciones preliminares sin comprometer la esencia del criterio judicial, dada la naturaleza predecible de las decisiones adoptadas en la generalidad de estos procesos.

En todos los supuestos, el propósito debe ser garantizar la supervisión humana efectiva, la transparencia algorítmica y la posibilidad de auditorías periódicas, con el fin de alinear el uso de estas tecnologías con principios éticos y jurídicos como la igualdad, la no discriminación, la privacidad de datos y la rendición de cuentas. Solo así se podrá diseñar un marco regulatorio que, además de adaptarse a la diversidad de aplicaciones, soluciones y tecnologías, preserve la legitimidad y la equidad en el ejercicio de la función jurisdiccional.

Adicionalmente, agregaría valor que las resoluciones que se dicten diferencien el tipo de usuario destinatario de las directrices, ya que tanto la implementación como los riesgos asociados al uso de IA e IA generativa varían según si el acceso y utilización de estas herramientas se realiza bajo un modelo By User (gestionado directamente por el usuario final) o By IT (administrado de forma centralizada por departamentos técnicos).

Es importante en este punto tener en cuenta que, a diferencia de los sistemas de IA tradicionales, cuya implementación solía depender exclusivamente de las áreas de informática para su configuración y control, la IA generativa introduce un paradigma en el que los usuarios finales —jueces, funcionarios o empleados— pueden emplear herramientas avanzadas de manera autónoma, automatizando tareas bajo su propia dirección sin mediación técnica.

En efecto, las experiencias recogidas en este piloto dan cuenta de una tendencia predominante hacia el modelo By User, gestionado directamente por el usuario final, lo que plantea desafíos y oportunidades específicas que también deben considerarse y que exigen estrategias regulatorias diferenciadas.

No obstante, es posible que la mayoría de los poderes judiciales avancen hacia tendencias híbridas, combinando sistemas inteligentes gestionados centralizadamente por departamentos técnicos (By IT) con soluciones que sean directamente administradas y, en algunos casos —sobre todo de cara a los agentes de IA—, incluso diseñadas por los propios usuarios (By User).

De esta manera, para los entornos de aplicación By IT, donde la implementación es gestionada centralmente, las directrices normativas deberían centrarse en lo relativo a las configuraciones de seguridad (como filtros de contenido o restricciones de acceso), auditorías institucionales y la integración con sistemas judiciales existentes, para asegurar que la IA cumpla con estándares de transparencia y trazabilidad definidos por el Poder Judicial.

En cambio, en los escenarios By User, donde los usuarios acceden individualmente a aplicaciones de IA generativa (ej. ChatGPT, Gemini, Copilot) o la IA generativa ya le llega embebida en herramientas cotidianas —como procesadores de texto o asistentes de correo—, la regulación debe priorizar la capacitación intensiva, la sensibilización sobre riesgos (como sesgos, alucinaciones o dependencia excesiva) y el fomento de un uso responsable.

Dado que el enfoque By User acelera la adopción de la IA generativa a la vez que puede dificultar su monitoreo exhaustivo, las directrices deben establecer principios que equilibren autonomía y supervisión, como un enfoque híbrido que combine políticas institucionales centralizadas con protocolos de autogestión regulada para los usuarios.

Esta diferenciación permitirá diseñar medidas de mitigación proporcionadas —desde la prevención de errores en tareas críticas hasta la optimización de procesos rutinarios—, garantizando que la integración de la IA en el ecosistema judicial sea segura y ética.

Aún con lo señalado, es importante tener presente que la evolución de los modelos de lenguaje grande requerirá revisiones periódicas para adaptarse al desarrollo continuo de la tecnología y sus implicancias en la justicia.

Co-organizaron



.UBA derecho



IALAB



JusMendoza.Lab



FACULTAD DE DERECHO Y CIENCIAS SOCIALES Y POLÍTICAS
UNIVERSIDAD NACIONAL DEL NORDESTE



LIDeSIA

LABORATORIO DE INVESTIGACIÓN Y DESARROLLO DE SOFTWARE E INTELIGENCIA ARTIFICIAL



DERECHO
Departamento de Derecho Judicial

Colaboraron

