El uso de inteligencia artificial y sus implicancias

TEMA 2: Ejercicio notarial en la era digital

Coordinadores: Santiago Francisco Oscar Scattolini y Federico Jorge Panero

Subcoordinadora: Cecilia García Puente

AUTORES:

MIRÓ, Mariano[i]

marianomiro77@gmail.com

CELENTANO, Daniela Micaela[ii]

danielacelentano@hotmail.com

PONENCIAS

- Sería prudente que para utilización de inteligencia artificial generativa (GAI) se utilicen mecanismos que mitiguen el grado de alucinaciones. Sin olvidar que somos los profesionales del derecho quienes, a fin de cuentas, debemos supervisar y cotejar los resultados obtenidos.
- La supervisión humana es fundamental para la detección de alucinaciones y prevención de sesgos en la utilización de IA.
- Sería propicio contar un repositorio de información jurídica, sistematizado, completo y actualizado. De esta manera, mejoraría la calidad de los datos de entrada y, habiendo ajustado la temperatura del modelo, disminuirían las alucinaciones. A su vez, permitiría a los profesionales una supervisión ágil de los resultados generados.
- El principio de reserva de humanidad se manifiesta como la necesidad de intervención humana efectiva, por lo que, viene a reafirmar que la inteligencia artificial es una herramienta tecnológica al servicio de los profesionales, pero no los sustituye. Por ello, es imperativo el deber de un uso responsable de la inteligencia artificial, a los fines de conservar la confianza de la sociedad puesta en los profesionales.
- La utilización de IA viene acompañada de riesgos asociados como la localización de los servidores en los que operan los modelos de lenguaje, lo que puede implicar una transferencia internacional de datos, así como cesión de datos personales para el entrenamiento y perfeccionamiento de los modelos. *Ergo*, resulta indispensable reflexionar sobre alternativas más seguras, como lo son, actualmente, recurrir a proveedores radicados en la Unión Europea, o bien, la utilización de modelos de lenguaje abiertos que puedan descargarse y ejecutarse localmente.

ÍNDICE

Sumario: Mediante el presente trabajo nos adentraremos en el mundo de la inteligencia artificial generativa (GAI), desde un aspecto técnico y jurídico. Primeramente, se analiza, el reciente fallo de la Sala II de la Cámara de Apelaciones en lo Civil y Comercial de Rosario en los autos *Giacomino, César A. y otros c/ Monserrat, Facundo D. y otros s/*

Daños y Perjuicios, donde un letrado presentó un escrito judicial en el cual citó jurisprudencia inexistente que, según manifestó, le había proporcionado una IA. Abordamos, también, las implicancias que puede tener el uso de IA en la labor notarial. Fundamentalmente, se plantean diferentes propuestas tendientes a mitigar las "alucinaciones" de la IA. Finalmente, se analizan los riesgos asociados a la utilización de IA en materia de protección de datos personales.

ANÁLISIS DE UN CASO RECIENTE

La reciente decisión de la Sala II de la Cámara de Apelaciones en lo Civil y Comercial de Rosario en los autos *Giacomino, César A. y otros c/ Monserrat, Facundo D. y otros s/ Daños y Perjuicios*[iii] pone de manifiesto una cuestión de directa relevancia para la práctica forense contemporánea: el uso de herramientas de **inteligencia artificial generativa** (GAI) por partes y profesionales del derecho puede dar lugar a la incorporación en la litis de citas y referencias que resultan inexistentes o imprecisas.

Lo dejaron en evidencia

En los citados autos elevados a Cámara para resolver los recursos de apelación y nulidad interpuestos por los demandados y la citada en garantía, el Tribunal se encontró imposibilitado de localizar dos citas jurisprudenciales vertidas por el letrado de la parte actora al contestar agravios de la recurrente y fue justamente ello lo que motivó el dictado de una **medida de mejor proveer** solicitándole que identifique la fuente de tales citas.

Fue allí cuando el letrado tuvo que retractarse y poner de manifiesto que dichas citas le fueron proporcionadas por una aplicación de IA que utilizó como fuente de búsqueda y admitió no haber verificado previamente la existencia de todas las referencias aportadas, agregando, que volcó en la contestación de agravios el resultado encontrado actuando de **buena fe**.

¿Dónde quedó la ética profesional y el deber de diligencia?

El Tribunal, en tal sentido, sostuvo que: "*Tal actitud, aún de buena fe, compromete la responsabilidad profesional del letrado no sólo ante el tribunal sino, especialmente, respecto de su cliente …".*

Seguidamente, el Tribunal citó a las **Normas de Ética Profesional del Colegio de Abogados de Rosario** donde se alude al ***deber de probidad*****,** por el cual el letrado "... *requiere además lealtad personal, veracidad, buena fe. Así, por ejemplo, no debe … formular afirmaciones o negaciones inexactas, efectuar en sus escritos citaciones tendenciosamente incompletas, aproximativas o contrarias a la verdad* ...".

El corriente año, el mencionado Colegio rosarino, dictó una serie de **Recomendaciones para la Implementación Ética y Responsable de Inteligencia Artificial Generativa en la Práctica Jurídica** donde se establecen los principios que deberán observar los profesionales allí matriculados con relación al uso de la Inteligencia Artificial Generativa (IAGEN), los lineamientos para una utilización ética y responsable de la misma, junto las responsabilidades profesionales. En el corpus citado se afirma que "*Los profesionales deben llevar adelante un* ***control humano exhaustivo****, apropiado, razonable y proporcional al tipo de utilización realizada y a las tareas y actos procesales involucrados, respecto de todas las respuestas proporcionadas por modelos y agentes basados en inteligencia artificial generativa. … La* ***responsabilidad del profesional*** *es inaudible respecto del contenido generado con apoyo en IA.*". En el mismo sentido, en cuanto a los **protocolos de supervisión humana**, se insiste en que "*es fundamental que quienes utilizan estas herramientas cuenten con un conocimiento profundo del tema o ámbito específico donde se aplican, y que únicamente recurran a ellas cuando estén en condiciones de realizar una evaluación crítica y una validación previa de los resultados, antes de incorporarlos a cualquier proceso decisorio*."

De igual manera, en el derecho comparado, la Asociación de Abogados en los Estados Unidos (American Bar Association - ABA), en julio de 2024, emitió la ***Formal Opinion 512*** donde se establece que los abogados que utilizan herramientas de inteligencia artificial generativa (GAI) deben cumplir con sus obligaciones éticas y revisar cuidadosamente los resultados obtenidos para evitar errores. Asimismo, se remarca el riesgo de obtener resultados inexactos y que los abogados deben ser conscientes de que las salidas pueden no siempre ser confiables, por ende, existe el deber de **revisar** los resultados; destacando que la formación y actualización en tecnologías es parte esencial de la competencia profesional requerida a los abogados.

En este sentido, en el caso en cuestión, el Tribunal ratifica el **deber de cotejar las fuentes** por el profesional interviniente, no pudiendo revelarse de la misma amparándose en la buena fe ya que la misma debe enmarcarse en el **deber de probidad y diligencia** que hacen a la deontología propia de la profesión.

El consentimiento informado no te libera

El Tribunal afirma que el consentimiento informado por parte del cliente no exonera al letrado de su deber de actuar con diligencia, "*pues no puede haber consentimiento válido alguno que releve a un letrado de su deber de cotejar las fuentes en las que basa sus posiciones jurídicas."*

En definitiva, la ética profesional y el deber de diligencia no admiten renuncia dado que hacen a la esencia de la profesión.

Las alucinaciones

El Tribunal, acertadamente, hace mención a las, ya mencionadas, "**alucinaciones de la IA**", advirtiendo acerca de las mismas que "*mientras los sistemas no se desarrollen al punto de no alucinar, es sumamente riesgoso y hasta temerario delegar la labor de búsqueda de jurisprudencia de soporte y luego volcarla sin cotejar la fuente, como aquí ocurrió".*

Como veremos, eliminar por completo las alucinaciones resulta difícil, pero bien podrían implementarse técnicas adecuadas ajustando el modelo de forma tal, tendiente a **reducir** estas alucinaciones.

En tal sentido, el Laboratorio Fintech y Legaltech del Colegio Público de la Abogacía de la Capital Federal (CPACF) publicó una **Guía para el uso de Inteligencia Artificial para Abogados***,* donde su primer objetivo es el uso responsable de la IA y la mitigación de riesgos, dentro del cual se encuentra el deber de "*Validar todo el contenido: Corrobore siempre la exactitud del antecedente informado por el sistema. Siempre revise lo que dice la IA, ya que puede alucinar y crear citas falsas. Es falta ética efectuar citas doctrinarias o jurisprudenciales inexistentes …*". El segundo objetivo de la Guía hace referencia, justamente, al control humano, resaltando el deber de una **supervisión humana permanente** y de **validar todo el contenido** obtenido por la IA dado que podría sugerirse

información ficticia; reafirmando, nuevamente, que la responsabilidad por el uso de la IA, en definitiva, es del letrado interviniente.

La sanción al letrado

Si bien el Tribunal no sanciona directamente al letrado, le hace un llamado de atención con fines preventivos. A su vez, resuelve oficiar al Colegio de Abogados de Rosario para que tome conocimiento de la problemática en cuestión a fin de que adopte las medidas que correspondan y, a su vez, difunda ante sus colegiados los riesgos que implica el uso de inteligencia artificial generativa sin supervisión de los resultados ofrecidos por la misma.

En este sentido, si bien, el Colegio de Abogados de Rosario, ya había elaborado recientemente las, ya mencionadas, Recomendaciones para la Implementación Ética y Responsable de Inteligencia Artificial Generativa en la Práctica Jurídica, quizás se torne necesario difundirlas con mayor ímpetu, o bien, aplicar las sanciones pertinentes a los colegiados, tendientes a que las mismas se efectivicen, evitando así el uso discrecional de los sistemas de IA.

La experiencia extranjera

Este episodio no es aislado: precedentes extranjeros —entre ellos el conocido episodio en torno a ***Mata v. Avianca***[iv] en Estados Unidos— muestran consecuencias análogas cuando las referencias generadas por sistemas de IA no son comprobadas. En aquel caso, la identificación de sentencias apócrifas derivó en la imposición de sanciones procesales a los responsables de la presentación.

Desde una perspectiva jurídico-deontológica, ambos supuestos confluyen en un punto central: la **buena fe** y la voluntad de utilizar herramientas modernas no exime al profesional de su deber de veracidad y diligencia. El uso de una herramienta tecnológica que facilita la confección de escritos no sustituye la obligación del abogado de corroborar la existencia, autenticidad y alcance de las fuentes invocadas. En el plano procesal, la introducción de referencias falsas o inexactas puede redundar en la pérdida de credibilidad del escrito, provocando desde oficios aclaratorios hasta sanciones disciplinarias o procesales según la gravedad y la intencionalidad.

En ambos casos se trataba de lo que se conoce como “Alucinaciones” de la IA.

¿Pero qué son las alucinaciones y qué podemos hacer para prevenirlas? Para entender esto, primero, debemos entender de qué se tratan estos modelos de IA conversacionales.

IMPLICANCIAS DE LA UTILIZACIÓN DE GAI EN LA LABOR NOTARIAL

Sin lugar a duda, la implementación de inteligencia artificial generativa (GAI), siempre que esté supervisada, trae grandes ventajas a nuestra labor cotidiana. Ello se materializa, por ejemplo, en la **automatización** y, consecuente, mayor rapidez en la ejecución de tareas que no sean complejas. En el día a día realizamos muchas tareas que requieren un análisis profundo y complejo, para lo cual es necesaria la intervención directa y personal del notario. A su vez, existen otras tareas diarias de menor envergadura que perfectamente podemos delegarlas en terceros que colaboren con nuestra labor (vg. empleados, destajistas, protocolistas, etcétera). La inteligencia artificial generativa viene a aportar en este punto, mediante la redacción de documentos, confección automatizada de formularios o declaraciones juradas, para que, en definitiva, la labor notarial y de nuestros colaboradores esté enfocada en un rol de supervisión.

En este sentido, se nos pone la “piel de gallina” cuando escuchamos frases como: *La inteligencia artificial viene a reemplazarnos y a reemplazar a las profesiones.* La realidad es que la IA no viene a reemplazarnos, siempre y cuando, nos apoderemos de ella y la implementemos para el perfeccionamiento de nuestra labor.

La IA produce y nosotros verificamos lo que produce. Esto hace que podamos enfocarnos mayor tiempo en labores que verdaderamente requieran de nuestros conocimientos y *expertise* sin tener que dedicarles tanto tiempo a labores que perfectamente pueden verse automatizadas.

Otro punto a favor que tiene la implementación de IA en nuestra labor es la **reducción de errores** propios del trabajo manual, vg. errores de tipeo. Hay errores que son propios de la labor humana, como los errores de tipeo. La realidad es que un error de tipeo puede no ser relevante, en algunos casos, pero en otros puede ser crucial si estamos frente a un error en elementos esenciales del acto. En este punto, mediante herramientas de IA, podemos mitigar esos tipos de errores.

Ahora bien, ¿qué **riesgos** puede implicar la utilización de GAI en nuestra labor diaria? Con relación a las ya mencionadas **alucinaciones**, si no tomamos los recaudos de supervisar los textos generados mediante GAI, puede suceder lo mismo que le ha sucedido al letrado en el caso analizado: la GAI alucinó y el letrado, al no cumplir con su deber de supervisar, generó un documento con contenido apócrifo. Con el aditamento de que, en nuestra función, al ser autores de un instrumento público, podríamos caer en falsedades ideológicas. Cuestión no menor a tener en cuenta.

Otro punto de inflexión, que, si bien no es objeto de análisis del presente trabajo, pero que no podemos dejar de mencionar, son los **sesgos**. Una definición de la RAE, en la acepción que consideramos más atinente, define al término *sesgo* de la siguiente manera: *Error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras*.[v]

En relación con ello, la directora general de la UNESCO, Audrey Azoulay, ha sostenido que: "*Cada día son más las personas que utilizan modelos de lenguaje en su trabajo, sus estudios y en casa. Estas nuevas aplicaciones de IA tienen el poder de moldear sutilmente las percepciones de millones de personas, por lo que incluso pequeños sesgos de género en su contenido pueden amplificar significativamente las desigualdades en el mundo real. Nuestra Organización pide a los gobiernos que desarrollen y apliquen marcos regulatorios claros, y a las empresas privadas que lleven a cabo un seguimiento y una evaluación continuos para detectar sesgos sistémicos, como se establece en la Recomendación de la UNESCO sobre la ética de la inteligencia artificial, adoptada por unanimidad por nuestros Estados miembros en noviembre de 2021*."

El uso irresponsable de la IA no genera riesgos en abstractos, sino violaciones concretas de derechos fundamentales. En este sentido, la supervisión humana es fundamental para la detección y prevención de sesgos en los sistemas de IA.

LA INTELIGENCIA ARTIFICIAL, EL PROCESAMIENTO DE LENGUAJE NATURAL Y EL MODELO TRANSFORMER

Una definición

Hay muchas definiciones de Inteligencia Artificial, pero la más común de encontrar es la siguiente: “La inteligencia artificial (IA) es una tecnología que permite a las computadoras y máquinas simular el aprendizaje humano, la comprensión, la resolución de problemas, la toma de decisiones, la creatividad y la autonomía.” [vi]

Un poco de historia

El estudio de la IA comenzó en la década de 1950 del siglo pasado con las primeras conferencias y estudios y de la mano del primer arquetipo de programación basado en reglas establecidas por el programador para que las computadoras pudieran seguir. La inteligencia de la máquina estaba limitada a cuán bien pudiera el programador representar todo el arquetipo de posibilidades.

Con el tiempo, y gracias a la evolución de los algoritmos y el aumento de la capacidad de cómputo de las mismas, nació el *Machine Learning* (Aprendizaje Automático), en adelante ML. En donde el programador ya no establece reglas para que la computadora siga, sino que le muestra una serie de datos (el *input)* y sus resultados (el *target)* y es la computadora la que haciendo uso de reglas de la estadística debe establecer las relaciones entre esos datos y sus resultados, todo ayudado por una función que determinada el error en la predicción y cuyo objetivo era minimizarlo.

Pero, a pesar del avance que supuso el nacimiento de los modelos de ML, éstos solamente podían analizar los datos de entrenamiento y no ir más allá de ellos. Los modelos clásicos de ML estaban limitados en su capacidad de poder entender la realidad completa de los datos con los cuales habían sido entrenados para poder mejorar la predicción. Para poder avanzar en ello fue necesario desarrollar un nuevo concepto de ML que utiliza Redes Neuronales, el *Deep Learning*.

Deep Learning es una sub-aplicación de ML que utiliza redes neuronales[vii] que permiten aprender diferentes representaciones de los datos de entrenamiento, aumentando así su capacidad para poder encontrar las relaciones entre los *inputs* y los *targets*.

Ver y hablar

Si bien las redes neuronales tienen múltiples aplicaciones, dos avances captaron especialmente la atención de la comunidad científica: permitir que las máquinas **vean** (por medio del procesamiento de imágenes) y que **hablen** (a través del procesamiento de lenguaje). En tareas de visión por computadora, los modelos clásicos de aprendizaje automático ya ofrecían soluciones, pero sus tasas de acierto eran claramente inferiores a las alcanzadas por las nuevas arquitecturas neuronales. Entre aproximadamente 2011 y 2015, las redes neuronales profundas basadas en convoluciones —conocidas como **redes neuronales convolucionales** (RNC o, por su sigla en inglés, **CNN**)— empezaron a imponerse en competencias internacionales de *computer vision*. Estas arquitecturas aprenden múltiples representaciones de la imagen (capas de características) y, en conjuntos de datos relevantes para la comunidad científica —como ImageNet— alcanzaron niveles de precisión que superaron ampliamente a los métodos previos, marcando un punto de inflexión: las computadoras adquirieron, en términos prácticos, la capacidad de “ver” con un nivel de desempeño hasta entonces inusitado.

El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) planteó un desafío distinto. El lenguaje humano es heterogéneo: vocabulario extenso, modismos, ambigüedad sintáctica y relaciones semánticas que se extienden a lo largo de frases y párrafos. Los primeros enfoques automáticos se basaban en modelos estadísticos de baja complejidad —por ejemplo, modelos de **bigramas** o **trigramas**, que capturan relaciones entre dos o tres palabras—; estos enfoques fueron útiles en tareas puntuales (como análisis de sentimiento o traducción simple), pero sus limitaciones eran patentes cuando se les exigía comprender contextos largos o matices semánticos complejos.

En la evolución del procesamiento del lenguaje natural se impusieron las redes neuronales recurrentes y, entre ellas, las *Long Short-Term Memory* (LSTM), una arquitectura diseñada para retener y propagar información relevante a lo largo de secuencias temporales mediante un estado interno y puertas de control. Gracias a ese

mecanismo, las LSTM superaron a los modelos estadísticos tradicionales en numerosas tareas de NLP (traducción automática, reconocimiento de voz, etiquetado), porque resultaban mejores captando dependencias de corto y mediano alcance. No obstante, las LSTM tienen límites prácticos en la gestión de dependencias muy largas: la información distante tiende a perder influencia frente a la más reciente. Así, en la oración "Messi es el mejor jugador del mundo, nació en Rosario y juega en el Inter de Miami", el modelo LSTM perdía el contexto y no podía relacionar muy bien los verbos "nació" y "juega" con el sujeto "Messi". Por último, su naturaleza secuencial dificulta la paralelización completa en GPU[viii], incrementando los tiempos de entrenamiento. Estas limitaciones fueron, en buena medida, el estímulo para el desarrollo posterior de arquitecturas.

Todo lo que precisas es atención

Todo cambió en 2017 con la publicación del artículo seminal **"Attention Is All You Need"** (Vaswani et al.). En él se presentó la arquitectura denominada **Transformer**, una forma de red neuronal diseñada específicamente para procesar secuencias de texto. A diferencia de las arquitecturas recurrentes, el Transformer se basa en el mecanismo de **auto-atención (self-attention)**, que permite al modelo valorar directamente la relación entre cualquier par de palabras de una misma secuencia —incluso si están muy separadas— y, con ello, captar dependencias a larga distancia.

Además, el uso de **multi-head attention** (atención en múltiples "cabezas") posibilita que el sistema aprenda simultáneamente distintas representaciones o relaciones del texto. Desde el punto de vista computacional, esta estructura facilita una mayor **paralelización** del entrenamiento en GPU/TPU —porque no está sujeta a un paso secuencial palabra a palabra—, lo que acelera el proceso en grandes conjuntos de datos. Cabe advertir, no obstante, que los Transformers también tienen costes y limitaciones (por ejemplo, la atención puede crecer en coste de memoria con la longitud de la secuencia), y, además, su rendimiento depende tanto del diseño arquitectural como de los recursos disponibles.

La aparición de esta nueva arquitectura fue una revolución en sí, lo que dio pie al desarrollo de nuevos modelos. Uno de los primeros en aparecer fue GPT-1 desarrollado por OpenAI y BERT desarrollado por Google ambos en el año 2018, aunque con enfoques diferentes. A medida que se fueron incorporando más datos de entrenamiento

y fue mejorando tanto la arquitectura, como los métodos de entrenamiento, estos modelos fueron también evolucionando.

Técnicamente, los LLM son modelos matemáticos y probabilísticos que aprenden, a partir de enormes *corpus* textuales, las correlaciones estadísticas entre unidades de lenguaje. Durante el entrenamiento, ajustan millones o miles de millones de parámetros —los llamados "pesos"— que codifican esas relaciones. Operar el modelo consiste en usar esos pesos para estimar la probabilidad del siguiente *token* (unidad mínima de texto: palabra, subpalabra o símbolo) y generar así secuencias coherentes. Esta capacidad predictiva explica el notable desempeño en tareas de generación y comprensión del lenguaje, pero también la naturaleza no infalible del resultado: **el modelo produce lo que es estadísticamente plausible, según sus datos de entrenamiento, no una verificación factual directa.**

¿Cómo puede un LLM realizar «operaciones matemáticas» sobre palabras? La respuesta está en un proceso previo de representación numérica: primero el texto se divide en *tokens*; a cada token se le asigna un vector numérico *(embedding)* de alta dimensión que codifica rasgos léxicos y estadísticos. Piensen en este vector como un mapa en el que cada palabra es un punto. Palabras con significado parecido quedan cerca; palabras distintas quedan lejos. Un *embedding* es ese mapa, pero en vez de coordenadas geográficas usa muchas dimensiones numéricas. Las operaciones matemáticas (distancia, producto punto) sirven para medir la "cercanía" semántica.

Esos *embeddings* iniciales son luego transformados por las capas del modelo —mediante operaciones de álgebra lineal (productos escalares, multiplicaciones por matrices y sumas ponderadas)— para generar representaciones contextualizadas: vectores que incorporan información sobre la relación del *token* con las demás palabras de la secuencia. Esas representaciones son las que permiten que el modelo calcule probabilidades, comparar similitudes y genere texto coherente.

GPT-3.5 y Chat GPT

A finales de 2022 OpenAI presentó una nueva familia de modelos GPT e introdujo una aplicación para que cualquier pudiera interactuar con ellos **Chat GPT**. Lo cual fue en sí

una revolución, por cuanto los modelos habían salido de su nicho de laboratorio y toda la humanidad comenzó a interactuar con ellos.

A Chat GPT se le fueron sumando más modelos de diferentes empresas como Gemini de Google, Claude Anthropic y Copilot de Microsoft, solo por nombrar algunos.

¿GPT y ChatGPT son lo mismo? La respuesta es: **no exactamente**.

GPT designa una familia de *modelos de lenguaje* (Large Language Models, LLM): Que como dijimos son sistemas matemáticos entrenados con grandes volúmenes de texto para predecir el siguiente token en una secuencia. Un modelo GPT incorpora el conocimiento aprendido hasta la fecha de su entrenamiento —lo que se conoce como *cut-off*— y, por sí solo, no consulta la web en tiempo real. Por esa razón, un modelo entrenado con datos anteriores a un hecho reciente (por ejemplo, un resultado deportivo ocurrido ayer) no podrá informarnos sobre ese suceso salvo que se le suministre la información por otro medio.

ChatGPT, en cambio, es una aplicación o interfaz que utiliza uno de esos modelos subyacentes y lo complementa con reglas, mensajes de sistema, gestión de historial de conversación y, en algunas configuraciones, herramientas adicionales (buscador web, plugins, acceso a bases jurídicas, sistemas de recuperación documental). Esas capas añadidas permiten que la aplicación —cuando están habilitadas— obtenga información actualizada o recupere documentos específicos antes de formular la respuesta.

Por tanto, la diferencia práctica es la siguiente: **el modelo** aporta la capacidad predictiva y lingüística; **la aplicación** aporta el contexto operativo, las instrucciones y, si corresponde, los mecanismos para comprobar o actualizar hechos. Además, las aplicaciones suelen enviar al modelo, no solo el *prompt* del usuario, sino también instrucciones de sistema y el historial de la conversación, lo que condiciona el resultado final.

En consecuencia, no debería sorprender que un mismo modelo produzca respuestas distintas según la aplicación que lo envuelva y las herramientas que ésta tenga habilitadas. De allí la importancia, en el ámbito jurídico, de conocer la implementación concreta (¿se usó navegación web? ¿se consultaron bases oficiales?) y de verificar siempre las fuentes cuando la respuesta se apoya en hechos o precedentes.

Las Alucinaciones. ¿Por qué los modelos inventan jurisprudencia inexistente?

Como se explicó, los llamados *chats* no son más que aplicaciones basadas en un **modelo de lenguaje de gran escala (LLM)**, que en esencia es una construcción matemática y probabilística diseñada para simular el lenguaje humano. Estos modelos no "saben" en sentido estricto: lo que hacen es calcular la probabilidad de que una secuencia de *tokens* (fragmentos de texto) siga a otra. Durante su entrenamiento, el modelo aprendió patrones de redacción y, entre ellos, los formatos habituales de las **citas jurisprudenciales**.

El problema es que, si no tiene acceso a una **base de datos confiable** que confirme la existencia real de un fallo, el modelo tiende a **extrapolar**: combina nombres de partes, salas y fechas siguiendo patrones estadísticos y genera así una cita que parece verosímil, pero que en realidad nunca existió. Desde la perspectiva del sistema, no hay diferencia entre reproducir una cita auténtica o inventar una falsa; lo único que importa es que la secuencia "suene" coherente.

La investigación actual busca **mitigar estas alucinaciones** mediante diversas técnicas (integración con bases verificadas, recuperación documental, retroalimentación humana). Sin embargo, los propios desarrolladores reconocen que el fenómeno no puede eliminarse por completo. Por ello, todas las plataformas incluyen advertencias expresas: las respuestas no son infalibles y deben ser verificadas siempre por el usuario antes de utilizarlas en un contexto profesional o jurídico.

MITIGANDO LAS ALUCINACIONES

Ingeniería de Prompt

Una primera estrategia es la llamada *ingeniería de prompt*. Consiste en diseñar mejor la instrucción que se le da al modelo, introduciendo restricciones claras: exigir que indique la fuente, que limite su respuesta a ciertos documentos o que trabaje únicamente con un corpus suministrado por el usuario. Así, por ejemplo, al consultar sobre normativa nacional puede pedirse que cite únicamente el **Boletín Oficial** o la base **Infoleg**, lo que reduce el margen de error. Sin embargo, esta técnica tiene límites: aunque el usuario pida fuentes, el modelo puede inventarlas si no tiene acceso a un repositorio confiable.

Sobre todo con las fuentes doctrinales y jurisprudenciales, las que no son fáciles de encontrar libremente en internet y que, a menudo, forman parte de un paquete de suscripción a un servicio de información jurídica.

Ajuste Fino del Modelo

Otra alternativa es el *fine-tuning* o ajuste fino. En este caso ya no se trabaja solo con la aplicación (como ChatGPT) sino directamente con el modelo base, a través de una API. El procedimiento consiste en reentrenar parcialmente el modelo con datos propios del dominio de interés (por ejemplo, textos jurídicos).

Técnicas modernas como LoRA o QLoRA permiten realizar ese ajuste de manera más eficiente, modificando solo parte de los parámetros del modelo.

También existen enfoques de **Instruct fine-tuning**, que buscan alinear al modelo con instrucciones humanas explícitas para que siga, de forma más fiel, órdenes en lenguaje natural, reduciendo respuestas irrelevantes o ambiguas.

Otra línea complementaria es el **Reinforcement Learning with Human Feedback (RLHF)** (Aprendizaje por Refuerzo con Retroalimentación Humana). En este esquema, entrenadores humanos califican múltiples respuestas del modelo según criterios de utilidad, veracidad y seguridad; esos juicios se utilizan luego para ajustar al modelo mediante técnicas de aprendizaje por refuerzo. El resultado es un sistema que no solo predice la próxima palabra, sino que está **alineado** con las expectativas humanas. Tanto el *instruct fine-tuning* como el RLHF son estrategias claves en los modelos comerciales actuales, porque aumentan la confiabilidad práctica y disminuyen, aunque no eliminan, la posibilidad de alucinaciones.

Todas estas técnicas de ajuste fino pueden disminuir alucinaciones porque orientan el modelo hacia un dominio específico, pero debe combinarse con una buena ingeniería de *prompt* y enfrenta el mismo problema de fondo: la necesidad de contar con fuentes confiables que alimenten ese entrenamiento.

Generación Aumentada por Recuperación - Fuente de los datos

Como hemos dicho, a pesar de que utilizar técnicas como la ingenieria de prompt y el ajuste fino del modelo aún persiste la problemática de una fuente de Datos. En derecho

persiste una **fragmentación de los datos**. Es decir, la dispersión de los datos en múltiples sistemas, plataformas y fuentes.

Uno de los principales desafíos de la fragmentación de datos es la dificultad de integrarlos y depurarlos para su uso, limitando así el potencial de la IA ya que genera barreras para acceder, integrar y analizar datos eficazmente.

La fragmentación de los datos, por un lado, dificulta su acceso, integración y análisis por parte de los sistemas de GAI; y, por otro lado, aumenta la complejidad de los profesionales del derecho de realizar una búsqueda exhaustiva al momento de supervisar los resultados obtenidos.

En tal sentido, deviene indispensable la existencia de infraestructuras que concentren fuentes jurídicas confiables; es decir, un **repositorio de información jurídica**, **sistematizado, completo y actualizado**.

Ese repositorio puede ser utilizado con la técnica considerada más eficaz en la práctica es la **Retrieval-Augmented Generation (RAG)**. Aquí el modelo se complementa con una **base de datos vectorial** que contiene la información de ese repositorio jurídico, previamente indexada en forma de vectores numéricos[ix], compatibles con los que maneja el modelo. Esto es, cuando el usuario formula una consulta, el sistema busca primero en esa base los fragmentos más relevantes desde el punto de vista semántico y los incorpora al *prompt* que recibe el modelo. El LLM genera la respuesta basándose en esos documentos concretos, lo que reduce la posibilidad de inventar contenido. Al tener acceso a ese repositorio permite al modelo tomar información fidedigna y actualizada sobre el tema de la consulta.

Aunque no elimina totalmente el riesgo, RAG mejora notablemente la confiabilidad de las respuestas, especialmente cuando la base vectorial está construida con fuentes oficiales o verificadas.

Reserva de humanidad

Por otro lado, la intervención humana para verificar la veracidad y existencia de las fuentes informadas por el LLM es crucial para evitar que las alucinaciones provoquen un daño.

El **Reglamento (UE) 2024/1689** dictado por el Parlamento Europeo el 13 de junio de 2024, en materia de IA, en su primer artículo establece que el objetivo del Reglamento es "…*promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable…"*. Por otro lado, el Reglamento (UE) 2016/679, más conocido como, el **Reglamento General de Protección de Datos (RGPD)** del mismo Parlamento, en su artículo 22, establece que las personas tienen derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado. Esto es a lo que se llama ***reserva de humanidad***.

En esta línea, la **UNESCO**[x] ha establecido **Diez principios básicos para un enfoque de la ética de la IA centrado en los Derechos Humanos**, dentro de los cuales se encuentra el principio de *Responsabilidad y rendición de cuentas*, por el cual, los sistemas de IA deben ser auditables y trazables y deben existir mecanismos de supervisión, evaluación de impacto, auditoría y diligencia debida. A su vez, el principio de *Supervisión y decisión humana* establece que "*Los Estados Miembros deberían velar por que siempre sea posible atribuir la responsabilidad ética y jurídica a personas físicas o a entidades jurídicas existentes*."

La reserva de humanidad implica que existen decisiones que no pueden ser totalmente automatizadas; las mismas deben ser **supervisadas** por un ser humano, preservando, de este modo, la intervención humana efectiva en las decisiones que afectan a los derechos fundamentales.

Este principio busca, por un lado, preservar valores humanos en la toma de decisiones, como son la empatía y la equidad, y, a su vez, garantizar que la IA se utilice como una herramienta tecnológica sin ser un sustituto del juicio humano.

Del mismo modo, en ámbito profesional, la reserva de humanidad se traduce en la necesidad de establecer límites en la toma de decisiones discrecionales por parte de la IA, quedando la supervisión a cargo de los profesionales que la utilicen. En este sentido,

se reitera el deber de diligencia que nos cabe a los profesionales en nuestro ejercicio, debiendo corroborar la información obtenida mediante sistemas de IA, promoviendo un uso responsable de los mismos.

USO DE LA IA Y PROTECCIÓN DE DATOS PERSONALES

El avance de la Inteligencia Artificial ha modificado profundamente la forma en que interactuamos con el entorno digital y la manera en que utilizamos las herramientas tecnológicas disponibles para el ejercicio profesional. En el **ámbito notarial,** se presenta una oportunidad especialmente atractiva: emplear sistemas de IA capaces de asistir en la redacción de documentos, procesando de manera automática los datos aportados por las partes, así como la información contenida en certificados e informes expedidos por registros y dependencias públicas. Esta posibilidad, sin embargo, viene acompañada de **riesgos** considerables que no pueden ser ignorados, especialmente en materia de protección de datos personales.

El principal problema radica en la **localización de los servidores** en los que operan los modelos de lenguaje más difundidos. La gran mayoría de las plataformas disponibles —sean de acceso gratuito o mediante suscripción— procesan la información en servidores situados en países como Estados Unidos o China. La República Argentina, a diferencia de lo que ocurre con la Unión Europea, no cuenta con tratados de protección de datos vigentes con estas jurisdicciones, lo que convierte a cualquier transferencia internacional de información personal hacia tales territorios en un acto jurídicamente riesgoso y, en muchos casos, incompatible con la legislación local.

La Ley 25.326 de Protección de Datos Personales regula expresamente estas cuestiones. El artículo 11 establece que la cesión de datos a terceros requiere en todos los casos el consentimiento previo, expreso e informado del titular. El artículo 12, por su parte, dispone que la **transferencia internacional de datos** sólo es válida cuando el país receptor garantiza un nivel adecuado de protección o cuando exista un tratado específico que lo habilite. La importancia de ambas disposiciones se hace evidente al examinar cómo funcionan los modelos actuales. Ninguno de los más utilizados, como ChatGPT, Gemini, Claude o DeepSeek, cuenta con servidores instalados en territorio argentino. Todos ellos, salvo el caso de Mistral que pertenece a una empresa francesa, procesan la

información en jurisdicciones que no ofrecen garantías reconocidas por nuestro ordenamiento jurídico. A ello se agrega que, en Estados Unidos, país en el que se concentran gran parte de los desarrollos, no existe aún una normativa federal integral en materia de protección de datos personales.

A esta situación se suma otro aspecto crítico: varios proveedores de servicios de inteligencia artificial, entre ellos OpenAI y Anthropic, expresan de manera explícita en sus políticas de uso que los datos ingresados por los usuarios a través de los prompts pueden ser reutilizados para el entrenamiento y perfeccionamiento de sus modelos. Desde la perspectiva jurídica, este uso constituye una verdadera **cesión de datos personales,** alcanzada por la prohibición del artículo 11 de la Ley 25.326. De este modo, el envío de información personal de nuestros requirentes a estas plataformas sin consentimiento informado implicaría una vulneración directa de la normativa vigente y, por consiguiente, un serio riesgo de responsabilidad profesional para el notario.

La problemática adquiere mayor gravedad si consideramos que en la actividad notarial no sólo se manejan datos personales básicos como nombre, domicilio o número de documento; sino que también, muchas veces, se manejan **datos sensibles** en los términos del artículo 7 de la Ley 25.326. Información vinculada a la salud, al estado familiar, a creencias religiosas o a la situación patrimonial puede aparecer en escrituras y actas. El tratamiento inadecuado de esta categoría especial de datos no sólo compromete el deber legal de protección, sino que también expone al notario a sanciones administrativas y, eventualmente, a reclamos judiciales por parte de los titulares de la información.

Frente a este panorama, resulta indispensable reflexionar sobre **alternativas** más seguras. Una primera opción consiste en recurrir a **proveedores radicados en la Unión Europea**, donde rige el Reglamento General de Protección de Datos Personales (RGPD), actualmente considerado el estándar más avanzado en la materia. Este es el caso de Mistral, una empresa francesa que, además de operar bajo las reglas del RGPD, ha establecido en sus términos de servicio que la información ingresada por los usuarios no se utilizará para el entrenamiento de sus modelos. La transferencia de datos hacia este tipo de jurisdicciones se encuentra amparada por el marco del Convenio 108 y del

Convenio 108+, ambos suscriptos por la Argentina y por Francia, lo que otorga un sustento jurídico sólido para este tipo de intercambios.

Una segunda opción, más exigente en términos técnicos, es la utilización de **modelos de lenguaje abiertos** que puedan **descargarse y ejecutarse localmente**. Estos modelos vienen en diferentes tamaños. Pero aquellos de hasta los siete mil millones de parámetros, pueden ser operados en equipos personales que cuenten con la potencia suficiente, como aquellos provistos de tarjetas gráficas de alto rendimiento que se utilizan normalmente para correr video juegos. Esta estrategia, aunque más costosa y demandante en lo tecnológico, tiene la ventaja de garantizar un control absoluto sobre la información procesada. Bajo este esquema, los datos nunca abandonan el entorno seguro del estudio notarial, lo que asegura el cumplimiento de los principios de privacidad desde el diseño y privacidad por defecto, ampliamente reconocidos en el derecho comparado como estándares de buenas prácticas en materia de protección de datos.

REFLEXIONES FINALES

Concluimos en que es imperiosa e impostergable la necesidad de capacitación por parte de los profesionales del derecho en el uso adecuado de los sistemas de inteligencia artificial.

La implementación de sistemas de IA, en todos los ámbitos, avanza a pasos agigantados, por ende, sería prudente que para su utilización se utilicen mecanismos que mitiguen el grado de alucinaciones, como los mencionados. Sin olvidar que somos los profesionales del derecho quienes, a fin de cuentas, debemos supervisar y cotejar los resultados obtenidos.

Sería propicio, además, contar un repositorio de información jurídica, sistematizado, completo y actualizado. De esta manera, mejoraría la calidad de los datos de entrada y, habiendo ajustado la temperatura del modelo, disminuirían las alucinaciones. A su vez, permitiría a los profesionales una supervisión ágil de los resultados generados.

El principio de reserva de humanidad se manifiesta como la necesidad de intervención humana efectiva, por lo que viene a reafirmar lo que tanto hemos defendido: que la inteligencia artificial es una herramienta tecnológica al servicio de los profesionales, pero

que no los sustituye. Ahora bien, implica de nuestro lado, a su vez, un uso responsable, a los fines de conservar la confianza de la sociedad en los profesionales.

La tentación de recurrir a estas herramientas no puede eclipsar la obligación legal y profesional de custodiar los datos de quienes acuden al notario en busca de seguridad y confianza. Ya sea a través de proveedores europeos que garanticen niveles adecuados de protección o mediante la implementación de modelos locales, el desafío consiste en integrar la innovación tecnológica sin sacrificar el respeto por los derechos fundamentales de las personas. Solo así la inteligencia artificial podrá convertirse en una aliada del notariado, en lugar de un riesgo para su práctica y su legitimidad social.

BIBLIOGRAFÍA CONSULTADA

- Ashish Vaswani et al. *Attention is All you Need*, 2017, Attention is All you Need
- Formal Opinion 512. American Bar Association (ABA). 2024.
- Giacomino, Cesar A. y Otros c/Monserrat, Facundo D. y Otros s/Daños y Perjuicios. Cámara de Apelaciones en lo Civil y Comercial de Rosario - Sala II. 2025.
- Guía para el uso de Inteligencia Artificial para Abogados. Laboratorio Fintech y Legaltech. Unidad de Innovación y Transformación Digital del CPACF. 2025
- Gregkihlstrom. La fragmentación de datos inhibe el potencial de la IA.
- Hochreiter & Schmidhuber 1997 y Vaswani et al. *Long Short-Term Memory*, 2017, https://www.researchgate.net/publication/13853244_Long_Short-Term_Memory
- Lerer, Ignacio Adrian. La Guía de IA del CPACF y el problema latente: La infraestructura digital que aún no tenemos. https://abogados.com.ar/la-guia-de-ia-del-cpacf-y-el-problema-latente-la-infraestructura-digital-que-aun-no-tenemos/37341
- MIT. Cuando la IA se equivoca: cómo abordar las alucinaciones y los sesgos de la IA.https://mitsloanedtech-mit-edu.translate.goog/ai/basics/addressing-ai-hallucinations-and-bias/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc
- Recomendaciones para la Implementación Ética y Responsable de Inteligencia Artificial Generativa en la Práctica Jurídica. Colegio de Abogados de Rosario. 2025
- Reglamento (UE) 2016/679 (RGPD) del Parlamento Europeo y del Consejo. Diario Oficial de la Unión Europea. 2016
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo. Diario Oficial de la Unión Europea. 2024.
- Roberto MATA vs AVIANCA INC, United States District Court Southern District of New York, 2023. Mata v. Avianca, Inc., No. 1:2022cv01461 - Document 54 (S.D.N.Y. 2023) :: Justia
- UNESCO. *Un enfoque de la IA basado en derechos humanos.* 2024. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics.
- AI Engineering. Chip Huyen. Editorial O'Reilly.
- Deep Learning with Python. Segunda Edición. François Chollet. Manning Publications Co.
- Agents. Julia Wiesinger et al. Google.

- Solving Domain-Specific Problems Using LLMs. Christopher Semturs et al. Google.
- Foundational Large Language Models & Text Generation. Mohammadamin Barektain et al. Google.
- Hallucination mitigation with Agentic AI NLP-Based Frameworks. Diego Gosmar et al.

---

[i] Escribano Público en ejercicio de C.A.B.A. Abogado por la Universidad Católica Argentina. Especialista en Ciencia de Datos del Instituto Tecnológico de Buenos Aires. Diplomado en Gobierno de Datos y Protección de Datos Personales de la UCEMA, actualmente cursando la Maestria en Data Management y Analytics del Instituto Tecnológico de Buenos Aires. Miembro de la Comisión de Innovación y Tecnologías del Colegio de Escribanos de C.A.B.A.

[ii] Escribana pública en ejercicio de C.A.B.A. Abogada por la Universidad de Buenos Aires, con orientación en Derecho notarial, registral e inmobiliario. Diplomada en Aspectos legales de los smart contracts por la Universidad de Salamanca. Diplomada en Economía digital y smart contacts por la Universidad de Buenos Aires. Carrera de Especialización en Derecho Informático, U.B.A., en curso. Co-directora de la "Revista Jurídica-Notarial, un enfoque tecnológico y disruptivo" de la editorial IJ Editores. Miembro de la Comisión de Innovación y Tecnologías y de la Comisión de Integración Profesional del Colegio de Escribanos de C.A.B.A.

[iii] País: República Argentina. Expediente: CUIJ 21-11893083-2. Fecha: 01-08-2025.

[iv] ROBERTO MATA vs AVIANCA INC, United States District Court Southern District of New York, 2023. Mata v. Avianca, Inc., No. 1:2022cv01461 - Document 54 (S.D.N.Y. 2023) :: Justia

[v] REAL ACADEMIA ESPAÑOLA: Diccionario de la lengua española, 23.ª ed., [versión 23.8 en línea]. <https://dle.rae.es> Fecha de la consulta: 18/09/2025..

[vi] ¿Qué es la inteligencia artificial o IA? De Cole Stryker y Eda Kavlakoglu ¿Qué es la inteligencia artificial o IA? | IBM

[vii] ¿Qué es una red neuronal? Sin autores identificados. ¿Qué es una red neuronal? | IBM

[viii] GPU: Graphics Processing Unit (Unidad de procesamiento gráfico). Graphics processing unit - Wikipedia

[ix] Un vector numérico es una lista ordenada de números que en el caso de los modelos LLM guardan información respecto de la palabra, su contexto y las relaciones.

[x] UNESCO, *Un enfoque de la IA basado en derechos humanos. 2024* https://www.unesco.org/es/artificial-intelligence/recommendation-ethics