

Criterios orientadores para un uso ético y responsable de la Inteligencia Artificial en la Abogacía

Documento elaborado por el Instituto de Derecho Informático del CPACF

Buenos Aires, julio de 2025



**COLEGIO PÚBLICO DE LA ABOGACÍA
DE LA CAPITAL FEDERAL**

1 - Introducción.

El Colegio Público de la Abogacía de la Capital Federal considera oportuno emitir una opinión que sirva de criterio orientador a los matriculados sobre el uso ético y responsable de la Inteligencia Artificial (IA) en el ejercicio profesional.

Estos lineamientos están destinados a los abogados que decidan integrar la IA en su práctica profesional, a efectos que puedan hacerlo de forma segura y competente, alineados con la Ley 23.187 y al Código de Ética del CPACF, y así poder lograr una representación legal competente mediante el cumplimiento de los requisitos, deberes y derechos fundamentales.

Se abordan conceptos básicos, objetivos, aplicaciones potenciales y riesgos derivados de implementar la IA sin la debida precaución.

2 - Pilares de uso.

El uso de cualquier herramienta tecnológica requiere determinadas competencias y conocimientos previos. Estos criterios se basan en cuatro grandes pilares:

i. Fomentar el uso ético y responsable de la IA:

Deben minimizarse riesgos y evitar errores, prejuicios o información falsa generada por IA, manteniendo el estándar de calidad en el ejercicio de la abogacía.

ii. Promover siempre el control humano adecuado y vedar la delegación de decisiones:

Es preciso señalar que la IA, aún la IA generativa (IAGen), dada su naturaleza, es proclive a generar contenido incorrecto o alucinaciones y producir resultados que contengan sesgos, susceptibles de generar discriminación. Estos aspectos adquieren una relevancia particular en el ámbito jurídico.

En consecuencia, la supervisión humana, provista de pensamiento crítico para analizar, corregir o modificar el contenido generado por la IAGen, constituye una actividad ineludible en el trabajo colaborativo con esta nueva herramienta.

iii. Protección de datos personales, deber de confidencialidad y secreto profesional:

Evitar cargar datos personales de nuestros clientes o terceros a una herramienta de IA. El abogado debe respetar rigurosamente todo secreto profesional y oponerse ante cualquier acción o actividad que atente contra este, salvo las excepciones previstas: a) Cuando el cliente así lo autorice; b) Si se tratare de su propia defensa.

iv. Promover la capacitación:

La IA cambia a un ritmo disruptivo. Ello conlleva la necesidad y responsabilidad de

actualización continua a fin de utilizarla con conocimiento y criterio. Es un deber inherente al ejercicio de la abogacía.

3 - Criterios orientadores.

I - Uso responsable y mitigación de riesgos.

- Sesgos inherentes: La IA es proclive a reflejar y amplificar los prejuicios existentes en los datos con los que fue entrenada. El abogado debe mantener una perspectiva crítica para identificar y corregir, dentro de la medida de lo humanamente posible, sesgos de género, socioeconómicos, raciales o de cualquier otra índole que puedan afectar la objetividad de su trabajo.
- Errores y alucinaciones: Tener presente que la IA es proclive a generar resultados con errores y/o con información falsa.
- Deber de información y transparencia: Mantenga una comunicación transparente con sus clientes y busque su autorización antes de compartir cualquier información personal.

II - Control humano y prohibición de delegación

- Validar todo el contenido: Siempre revise exhaustivamente lo que produce la IA, ya que puede generar errores, alucinar y crear citas falsas. Por ejemplo, efectuar citas doctrinarias o jurisprudenciales inexistentes, o exponerlas en forma tal que falseen la opinión o el fallo invocados, o realizar falsas transcripciones de resoluciones judiciales o escritos del contrario.
- Calidad y actualidad de los datos de entrenamiento: Los modelos de IA, han sido entrenados por medio de legislación, jurisprudencia y doctrina hasta una fecha determinada, y no cuentan con información posterior. El abogado debe ser consciente de aquella limitación.
- No delegar decisiones: La IA puede ayudar al abogado, pero nunca debe sustituir su criterio humano.
- La responsabilidad es del abogado: Todo escrito o presentación que lleve la firma del abogado es de su única y exclusiva responsabilidad del profesional, independientemente de las herramientas utilizadas para su elaboración.

III - Protección de datos personales y deber de confidencialidad.

- Principio de minimización de datos: Como principio general, evite subir información

con datos personales, excepto que fuera estrictamente necesario para la tarea y que contara con el consentimiento expreso e informado de su cliente. Analice si la plataforma utilizada presenta nivel de seguridad y confidencialidad adecuado.

- Anonimizar información: Procure quitar o reemplazar nombres, documentos filiatorios u otra información personal.
- Evaluar las políticas de privacidad: No todas las herramientas protegen de igual modo la información cargada. Lea detenidamente los términos y condiciones de la herramienta de IA que use, ya que no todas ofrecen un nivel de protección de datos adecuado.
- Formarse en buenas prácticas: Todos los abogados deben contar con nociones fundamentales sobre protección de datos personales, seguridad digital y seguridad de la información.

IV - Formación y capacitación continua.

- Asistir a capacitaciones prácticas: Infórmese sobre cómo redactar mejores prompts y cómo detectar sesgos o errores.
- Actualizarse constantemente: La IA avanza a una velocidad exponencial; quienes no se formen se exponen a una creciente brecha de conocimientos y competitividad.

4 - Prompting.

I - Concepto general de prompting.

El prompting es el proceso mediante el cual el abogado guía a un modelo de IA para obtener resultados precisos, relevantes y útiles. La calidad de la respuesta de la IA es directamente proporcional a la calidad de la instrucción recibida. Un prompt bien construido es la diferencia entre un resultado genérico respecto de uno que acelere y enriquezca el trabajo del abogado.

II - Elementos esenciales de un prompt jurídico efectivo:

Para maximizar la eficacia, una instrucción bien diseñada debería contener, en la medida de lo posible, los siguientes elementos:

- Asignación de rol (persona): Instruya a la IA para que adopte una identidad específica. Esto enfoca su base de conocimiento y su estilo de respuesta. Ejemplo: "Actuar como un abogado especialista en derecho procesal civil y comercial, con 20 años de experiencia en litigios complejos".

- Contexto detallado: Proporcione el trasfondo fáctico y jurídico necesario. Explique la situación del cliente, la etapa procesal, la jurisdicción aplicable y el objetivo estratégico. Cuanto más rico el contexto, más pertinente será la respuesta.
- Tarea específica y delimitada: Sea explícito sobre la acción que la IA debe realizar.
- Tono y estilo: Defina el registro lingüístico deseado. Puede ser formal, técnico, persuasivo, conciliador o explicativo para un cliente no versado en derecho.
- Formato de salida: Indique cómo desea recibir la información para que sea directamente utilizable. Ejemplo: "Presentar la información en una tabla comparativa de tres columnas", "Elaborar un resumen ejecutivo de no más de 200 palabras", "Generar una lista de puntos clave".

III - El proceso iterativo: dialogar para refinar

El prompting no es un acto único, sino un diálogo. La primera respuesta de la IA es un punto de partida. Resulta así fundamental:

- Analizar y criticar el resultado: Evaluar la respuesta inicial, identificar imprecisiones, omisiones o áreas de mejora.
- Refinar el Prompt: Utilice la respuesta inicial para hacer preguntas de seguimiento más específicas. Corrija a la IA si comete un error. Pídale que profundice en un punto determinado, que considere una perspectiva alternativa o que reformule una sección.

IV - Técnicas avanzadas para la práctica jurídica.

- Proveer una base de conocimiento (prompting en contexto): Para obtener resultados de máxima precisión y evitar alucinaciones, la técnica más segura es adjuntar o pegar el material sobre el cual la IA debe trabajar. Suba los extractos pertinentes de la demanda, la contestación, la jurisprudencia relevante o la doctrina aplicable y ordene a la IA que base su respuesta exclusivamente en esa información.
- Exigir verificación y cita de fuentes: Si utiliza un modelo con acceso a internet, exija que cite las fuentes de sus afirmaciones (leyes, decretos, fallos). Verifique sistemáticamente cada fuente proporcionada, ya que pueden ser imprecisas o estar desactualizadas.
- Instruir en contra de la invención: Incluir una directiva explícita en su prompt para mitigar la fabricación de datos. Ejemplo: "No inventes ni supongas información. Si no conoces la respuesta o la normativa aplicable, indicarlo expresamente".
- Evitar ambigüedades: La IA interpreta las instrucciones. Las contradicciones o imprecisiones en el prompt conducirán a resultados genéricos o erróneos. Revise su instrucción para asegurarse de que sea clara, coherente y unívoca.

5 - Conclusiones.

El uso de la IA en el marco del ejercicio profesional siguiendo estos lineamientos ayudará a promover su aplicación ética, responsable y transparente y así evitar poner en riesgo su trabajo profesional, los intereses de su cliente y su reputación.

Es importante recordar que:

- La IA generativa puede ser una aliada para el mundo legal. Pero como toda herramienta tecnológica necesita que su uso esté guiado por reglas claras, criterio profesional y una mirada crítica.
- La IA no reemplaza el criterio jurídico del abogado ni su conocimiento de la ley: puede fallar y generar información falsa, por lo que requiere que se supervise rigurosamente sus resultados.
- La IA puede consolidar desigualdades producidas por sus sesgos, en caso de no ser utilizada con perspectiva crítica y atención a los derechos fundamentales involucrados.

6 - Glosario

A

Algoritmo de IA:

Conjunto de reglas o instrucciones que una computadora sigue para aprender de datos, resolver un problema o realizar una tarea y tomar decisiones.

Algoritmo de IA verde:

Es un método de programación que busca reducir el impacto ambiental de las aplicaciones informáticas, así como su uso para encontrar soluciones sostenibles. Se diferencian de los rojos —también llamados negros en relación al color característico que se asocia al CO₂— en que optimizan el uso de recursos de la computadora como la memoria, el procesador, la batería y el ancho de banda, minimizando así la huella de carbono de los dispositivos o redes de comunicación.

Alucinación:

Se refiere a aquellos casos en que el modelo de lenguaje genera resultados que pueden ser incorrectos, inexactos o inventados. Esto ocurre cuando el modelo de IA no está seguro del contexto, depende demasiado de sus datos de entrenamiento o carece de una comprensión adecuada del tema que se le presenta. La información producida incorrecta, inventada o inexacta, se presenta al usuario como si fuera precisa o factual.

Análisis de Datos:

Proceso de examinar, limpiar y modelar grandes conjuntos de información con el objetivo de extraer información útil para la toma de decisiones.

Anclaje:

Es el proceso de fijar en los sistemas de IA experiencias, conocimientos o datos del mundo real, o sea basarse en una información inicial o en ciertos valores. El objetivo es mejorar la comprensión del mundo por parte de la IA, para que pueda interpretar y responder efectivamente a las entradas, consultas y tareas de los usuarios.

Apilamiento (staking):

Es una técnica en IA que combina múltiples algoritmos para mejorar el rendimiento general y la consecuente obtención de mejores resultados. Compensa las debilidades de cada modelo y logra una salida más precisa y robusta en diversas aplicaciones, como el reconocimiento de imágenes y el procesamiento del lenguaje natural.

Aprendizaje automático (machine learning, ML):

Es una rama de la inteligencia artificial que permite a las computadoras aprender y mejorar a partir de datos sin ser programadas explícitamente. Utiliza algoritmos para identificar patrones y hacer predicciones o decisiones basadas en esos datos. Existen diferentes tipos de aprendizaje, como supervisado, no supervisado y por refuerzo. Su aplicación abarca desde el reconocimiento de imágenes hasta la predicción de tendencias en datos.

Aprendizaje Supervisado por humanos:

Es un subtipo del aprendizaje automático (machine learning). Implica entrenar un modelo con datos etiquetados, donde se conoce la respuesta correcta.

Aprendizaje no supervisado:

Es un subtipo del aprendizaje automático (machine learning). El modelo trabaja con datos sin etiquetas, buscando patrones o agrupaciones.

Aprendizaje no supervisado por Refuerzo:

Es un subtipo del aprendizaje automático (machine learning). Se basa en la interacción con un entorno, donde el agente aprende a tomar decisiones mediante recompensas y penalizaciones.

Aprendizaje colectivo:

Es un enfoque de entrenamiento de IA que aprovecha habilidades y conocimientos diversos a través de múltiples modelos para lograr una inteligencia más poderosa y robusta.

Aprendizaje de disparo cero o entrenamiento cero (Zero Shot Learning - ZSL):

Es un enfoque en el aprendizaje automático donde un modelo puede reconocer clases no vistas durante su entrenamiento. En lugar de depender de ejemplos etiquetados, ZSL utiliza información adicional, como descripciones o características de las clases, para inferir y clasificar nuevas categorías.

Aprendizaje potenciado:

Es una técnica en aprendizaje automático que combina múltiples modelos débiles para crear un modelo más robusto y preciso. A través de un proceso iterativo, se ajustan los modelos sucesivos para corregir errores de los anteriores, enfocándose en los ejemplos difíciles.

Aprendizaje Profundo (Deep Learning, DL):

Es una subrama del aprendizaje automático que utiliza redes neuronales artificiales con múltiples capas para modelar patrones complejos en grandes volúmenes de datos.

Estas redes pueden aprender representaciones jerárquicas, lo que les permite realizar tareas como reconocimiento de voz, imagen y procesamiento del lenguaje natural. Su capacidad para automatizar la extracción de características las hace especialmente efectivas en diversas aplicaciones.

Aproximación Bayesiana:

Es un enfoque estadístico que utiliza el teorema de Bayes para actualizar la probabilidad de una hipótesis a medida que se obtienen nuevos datos. Este método permite integrar la incertidumbre y realizar inferencias sobre parámetros o modelos basándose en evidencia previa y observaciones actuales. Es ampliamente utilizado en aprendizaje automático, estadística y análisis de datos para modelar situaciones donde la información es incompleta o incierta.

Asistente conversacional basado en IA generativa:

Es un sistema diseñado para interactuar con los usuarios a través de conversaciones naturales, utilizando modelos de lenguaje avanzados. Estos asistentes pueden entender y generar texto de manera coherente y contextual, respondiendo a preguntas y llevando a cabo tareas. Su capacidad para aprender de grandes volúmenes de datos les permite adaptarse y mejorar continuamente en la interacción con los usuarios.

Auditoría de algoritmos:

Es un proceso que evalúa y verifica el funcionamiento, la transparencia y la equidad de un algoritmo. Este análisis busca identificar sesgos, errores y posibles impactos negativos en los resultados generados por el algoritmo. La auditoría es fundamental para garantizar la responsabilidad y la confianza en sistemas automatizados, especialmente en aplicaciones sensibles como la contratación, la justicia y la salud.

Automatización:

La automatización se refiere al uso de tecnología para realizar distintas tareas con mínima intervención humana.

B

Base de datos:

Un sistema organizado que almacena información de manera estructurada, permitiendo su fácil acceso y gestión. Se utiliza para guardar datos que pueden ser consultados y actualizados de forma eficiente, facilitando la toma de decisiones y el análisis de información en diversas aplicaciones.

Big Data:

Se refiere a conjuntos de datos extremadamente grandes y complejos que no pueden ser gestionados ni analizados con métodos tradicionales. Estos datos provienen de diversas fuentes, como redes sociales, sensores y transacciones en línea, y requieren herramientas especializadas para su procesamiento. Permite descubrir patrones y tendencias valiosas que pueden informar decisiones estratégicas en negocios y otras áreas.

Blockchain e Inteligencia Artificial:

Refiere a la utilización de la tecnología de cadena de bloques para mejorar la seguridad, transparencia y eficiencia de los sistemas de inteligencia artificial. Blockchain permite un registro inmutable de datos, lo que puede ayudar a garantizar la integridad de la información utilizada por los modelos de IA. A su vez, la IA puede optimizar procesos en redes blockchain, como la detección de fraudes y la gestión de contratos inteligentes.

C

Cadena de Pensamiento (Chain of Thought, CoT):

Consiste en estructurar las instrucciones de forma tal que el modelo de lenguaje simule un proceso de razonamiento paso a paso antes de llegar a una conclusión. Esta estrategia es especialmente eficaz para tareas complejas que requieren análisis lógico, inferencia jurídica, interpretación normativa o evaluación de múltiples factores, como suele ocurrir en el ámbito del Derecho.

En lugar de pedir al modelo una respuesta directa, CoT descompone el problema en pasos intermedios, guiando al LLM para que construya su razonamiento de forma ordenada y transparente. Esto no solo mejora la precisión del resultado, sino que permite auditar el proceso lógico del modelo, algo fundamental en entornos sensibles como el jurídico.

Capacidad de control:

Es la habilidad de supervisar, ajustar y gestionar los sistemas de inteligencia artificial para asegurar que funcionen de manera correcta y ética. Esto implica establecer límites, monitorizar decisiones y realizar intervenciones cuando sea necesario, garantizando que los resultados sean confiables y alineados con las normas y expectativas sociales.

Ciencia de Datos:

Es un campo interdisciplinario que utiliza métodos estadísticos, algoritmos y técnicas de programación para analizar y extraer información valiosa de grandes conjuntos de datos. La ciencia de datos combina conocimientos de matemáticas, informática y dominio del problema para transformar datos en insights que puedan guiar la toma de decisiones en diversas industrias.

Chatbot:

Es un programa de software diseñado para simular conversaciones humanas a través de texto o voz. Utiliza inteligencia artificial y procesamiento de lenguaje natural para entender preguntas y proporcionar respuestas relevantes. Los chatbots se emplean en atención al cliente, asistencia en línea y aplicaciones interactivas, mejorando la eficiencia y accesibilidad de la comunicación con usuarios.

ChatGPT:

Es un modelo de lenguaje desarrollado por OpenAI que utiliza inteligencia artificial para generar texto de manera coherente y relevante en respuesta a las consultas de los usuarios. Capaz de entender y producir lenguaje natural, ChatGPT se emplea en diversas aplicaciones, como asistencia virtual, generación de contenido y educación, proporcionando interacciones más fluidas y contextuales.

Ciberseguridad en IA:

Se refiere a las prácticas y tecnologías diseñadas para proteger sistemas de inteligencia artificial de amenazas y ataques cibernéticos. Esto incluye asegurar la integridad de los datos utilizados para entrenar modelos, prevenir el acceso no autorizado y mitigar riesgos asociados con la manipulación de algoritmos. La ciberseguridad en IA es crucial para garantizar la confianza y la seguridad en aplicaciones que dependen de estos sistemas.

Clasificación:

Es un proceso de análisis de datos que implica asignar categorías o etiquetas a elementos basándose en sus características. En el contexto de inteligencia artificial, se utilizan algoritmos para entrenar modelos que pueden identificar y categorizar nuevos datos, facilitando tareas como el reconocimiento de imágenes, la detección de fraudes y el filtrado de spam. La clasificación ayuda a organizar información y tomar decisiones informadas.

Convolución:

Es una operación matemática utilizada en el procesamiento de señales y en redes neuronales, especialmente en las redes neuronales convolucionales (CNN). En este contexto, la convolución permite extraer características de datos, como imágenes, al aplicar filtros que destacan patrones y texturas. Esta técnica es fundamental para tareas como el reconocimiento de imágenes y la visión por computadora, mejorando la capacidad de los modelos de IA para interpretar datos visuales.

Conciencia algorítmica (Algorithmic Awareness):

Se refiere a la comprensión de cómo los algoritmos toman decisiones y el impacto que

estas decisiones pueden tener en la sociedad. Implica reconocer los sesgos potenciales y las limitaciones de los sistemas algorítmicos, así como la responsabilidad de quienes los diseñan y utilizan. Fomentar la conciencia algorítmica es crucial para asegurar que la tecnología se implemente de manera ética y justa.

Confiabilidad (Accuracy):

Es una medida que indica cuán precisas son las predicciones de un modelo de inteligencia artificial en comparación con los resultados reales. La confiabilidad es esencial para evaluar el rendimiento de los sistemas de IA, ya que una alta precisión garantiza que las decisiones basadas en estos modelos sean más seguras y efectivas.

D

Datamining (Análisis y minería de datos):

Es el proceso de explorar grandes conjuntos de datos para descubrir patrones, tendencias y relaciones significativas. Utiliza técnicas estadísticas y algoritmos de aprendizaje automático para transformar datos en información útil, que puede informar decisiones estratégicas. La minería de datos se aplica en diversas áreas, como marketing, finanzas y salud, para identificar oportunidades y optimizar procesos.

Datos:

Son hechos, cifras o información que se recopilan y pueden ser analizados para extraer significado. Los datos pueden estar en diferentes formatos, como números, texto o imágenes, y son la base para el análisis y la toma de decisiones en diversas disciplinas. Su correcta gestión y análisis son fundamentales para generar insights y apoyar estrategias en diferentes contextos.

Datos Abiertos:

Son conjuntos de datos que están disponibles al público de manera libre y accesible, permitiendo su uso, reutilización y redistribución sin restricciones. Pueden provenir de gobiernos, organizaciones y otras entidades, y se presentan en formatos que facilitan su análisis. Fomentan la transparencia, la innovación y la colaboración en diversas áreas, permitiendo que investigadores, desarrolladores y ciudadanos los utilicen para generar nuevos conocimientos y soluciones.

Datos de Entrenamiento:

Son conjuntos de datos utilizados para enseñar a un modelo de inteligencia artificial a realizar tareas específicas. Estos datos incluyen ejemplos representativos que ayudan al modelo a identificar patrones y aprender a hacer predicciones o clasificaciones. La calidad y la diversidad de los datos de entrenamiento son cruciales para el rendimiento del modelo, ya que influyen en su capacidad para generalizar a nuevas situaciones.

Datos no estructurados:

Son información que no sigue un formato predefinido ni se organiza en bases de datos

tradicionales. Ejemplos incluyen texto libre, correos electrónicos, imágenes y videos. Debido a su naturaleza caótica, estos datos requieren técnicas avanzadas de análisis y procesamiento, como el procesamiento de lenguaje natural y el aprendizaje automático, para extraer insights útiles. Son valiosos en la toma de decisiones, pero presentan desafíos en su gestión y análisis.

Datos Personales:

En Argentina, de acuerdo al artículo 2 de la Ley de Protección de Datos Personales (Ley 23.526) se entiende por "datos personales" a la información de cualquier tipo referida a personas físicas o de existencia ideal determinadas o determinables.

A nivel internacional, el Reglamento General de Protección de Datos (RGPD) que rige en la Unión Europea, define como "datos personales" es toda información sobre una persona física identificada o identifiable («el interesado»); se considerará persona física identifiable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona.

Datos de validación:

Son conjuntos de datos utilizados para evaluar el rendimiento de un modelo de inteligencia artificial después de su entrenamiento. Su función es verificar si el modelo generaliza correctamente y hace predicciones precisas en datos que no ha visto antes. Los datos de validación ayudan a ajustar y optimizar el modelo, asegurando que sea robusto y efectivo antes de su implementación en situaciones reales.

Deepfake:

Es una técnica que utiliza inteligencia artificial para crear contenido multimedia falso que parece real, como videos o audios. Esta tecnología puede alterar o superponer la imagen y voz de una persona en material existente, generando resultados muy convincentes. Aunque tiene aplicaciones creativas y educativas, los deepfakes también presentan riesgos significativos, como la difusión de desinformación y la manipulación malintencionada de la opinión pública.

DeepSeek R1:

Es un modelo avanzado de inteligencia artificial que ha llamado la atención por su capacidad para resolver problemas complejos a través del razonamiento lógico y el análisis de datos

DeepSeek-V3:

Es un asistente de inteligencia artificial diseñado para proporcionar respuestas precisas, información útil y apoyo en diversas tareas, desde consultas simples hasta análisis complejos.

Derechos digitales:

Son los derechos que protegen a los individuos en el entorno digital, abarcando aspectos

como la privacidad, la libertad de expresión y el acceso a la información. Estos derechos buscan garantizar que las personas puedan utilizar la tecnología de manera segura y justa, defendiendo su información personal y su capacidad para participar en la sociedad digital. La protección de los derechos digitales es fundamental en un mundo cada vez más conectado y dependiente de la tecnología.

Descenso de Gradiente:

Es un algoritmo de optimización utilizado en el entrenamiento de modelos de inteligencia artificial, especialmente en redes neuronales. Su objetivo es minimizar la función de pérdida ajustando los parámetros del modelo mediante la actualización en la dirección opuesta al gradiente de la función. Este proceso iterativo ayuda a encontrar los valores óptimos que mejoran la precisión del modelo en sus predicciones, contribuyendo a su eficacia y rendimiento general.

E

Encadenamiento de modelos:

Es una técnica que consiste en combinar múltiples modelos de inteligencia artificial para mejorar el rendimiento general de un sistema. Al encadenar modelos, cada uno puede especializarse en diferentes aspectos de una tarea, permitiendo que la salida de un modelo se utilice como entrada para otro. Este enfoque puede aumentar la precisión, la robustez y la capacidad de generalización del sistema, abordando problemas complejos de manera más efectiva.

Enriquecimiento de datos:

Es el proceso de mejorar la calidad y el valor de un conjunto de datos al agregar información adicional o relevante. Esto puede incluir la integración de datos de diferentes fuentes, la adición de etiquetas o la transformación de datos para hacerlos más útiles. El enriquecimiento de datos permite obtener insights más profundos y precisos, mejorando la toma de decisiones y la efectividad de los modelos de inteligencia artificial.

Espacio latente:

Es una representación abstracta de datos en la que se capturan las características más relevantes y las relaciones subyacentes. En el contexto de modelos generativos, como las redes neuronales, el espacio latente permite a los algoritmos aprender y organizar información de manera que facilite la generación de nuevos datos similares. Este concepto es fundamental para tareas como la generación de imágenes, texto y otras formas de contenido, ya que ayuda a entender la estructura de los datos.

Ética de los datos:

Se refiere a las normas y principios que guían el uso responsable y justo de los datos en la investigación y la inteligencia artificial. Esto incluye consideraciones sobre la privacidad, la transparencia, el consentimiento informado y la equidad. La ética de los datos es crucial para proteger los derechos de los individuos y garantizar que los datos

se utilicen de manera que beneficien a la sociedad, evitando abusos y discriminación en los procesos de análisis y toma de decisiones.

Etiquetado de Datos:

Es el proceso de asignar etiquetas o categorías a conjuntos de datos para facilitar su análisis y uso en el entrenamiento de modelos de inteligencia artificial. Este proceso es esencial para enseñar a los modelos a reconocer patrones y realizar tareas específicas, como la clasificación de imágenes o el análisis de texto. Un etiquetado preciso y coherente mejora la calidad del modelo y su capacidad para generalizar a nuevos datos.

Explicabilidad (XAI):

Se refiere a la capacidad de entender y explicar cómo y por qué un modelo de inteligencia artificial toma decisiones específicas. Es fundamental para generar confianza en los sistemas de IA, especialmente en aplicaciones críticas como la medicina y la justicia. Al proporcionar claridad sobre el funcionamiento del modelo, se permite a los usuarios evaluar su lógica y evaluar la validez de sus resultados, promoviendo un uso más responsable y ético de la inteligencia artificial.

Extensibilidad:

Es la capacidad de un sistema o modelo de inteligencia artificial para adaptarse y expandirse con el tiempo, incorporando nuevas funcionalidades o mejoras sin requerir una reestructuración completa. Permite a los desarrolladores integrar nuevas tecnologías, algoritmos o datos, facilitando la evolución del sistema y su ajuste a las cambiantes necesidades del usuario o del entorno. Este atributo es fundamental para mantener la relevancia y eficacia de las aplicaciones de IA a largo plazo.

Extracción:

Es el proceso de obtener información relevante y útil de grandes volúmenes de datos. En el contexto de inteligencia artificial y ciencia de datos, la extracción implica identificar y recolectar datos significativos que pueden ser analizados para revelar patrones, tendencias y relaciones. Este paso es crucial para preparar los datos para su análisis posterior, asegurando que se utilicen las variables adecuadas para alimentar modelos y tomar decisiones informadas.

F

Few-shot prompting:

Cuando se requiere un resultado más preciso, se agrega al prompt múltiples ejemplos. Esto permitirá a la IA comprender mejor el patrón subyacente y la tarea encomendada: Por ejemplo: "Determina la rama del derecho a la que pertenece la siguiente norma: 'Ley que regula el matrimonio igualitario' - 'Derecho de familia'; 'Artículo del Código Penal sobre el hurto' - 'Derecho penal'; 'Ley de defensa del consumidor' - 'Derecho del consumidor'. Ahora determina la rama del derecho de: 'Artículo de la Constitución Nacional sobre la libertad de expresión'"

Función de Activación:

Es una fórmula matemática utilizada en redes neuronales para determinar si una neurona debe activarse o no, es decir, si debe enviar su señal de salida. La función de activación introduce no linealidades en el modelo, permitiendo que la red aprenda patrones complejos en los datos. Existen diferentes tipos de funciones de activación, como la sigmoide, ReLU (Rectified Linear Unit) y tanh, cada una con sus propias características y aplicaciones en el entrenamiento de modelos de inteligencia artificial.

G

Generalización:

Es la capacidad de un modelo de inteligencia artificial para aplicar lo aprendido durante su entrenamiento a datos nuevos y no vistos. Un modelo bien generalizado es capaz de hacer predicciones precisas en situaciones diferentes a las que fue entrenado, evitando el sobreajuste, que ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento y pierde capacidad de adaptación. La generalización es crucial para la efectividad de un modelo en aplicaciones del mundo real.

Generalización de débil a fuerte:

Se refiere al proceso en el que un modelo de inteligencia artificial comienza con habilidades limitadas para hacer predicciones y, a medida que se entrena con más datos y se ajusta, mejora su capacidad para generalizar y realizar predicciones más precisas en una variedad de situaciones. Este concepto implica que el modelo evoluciona de una comprensión básica a una comprensión más robusta y compleja, lo que permite una mejor adaptación a nuevos desafíos y contextos.

Generación de Lenguaje Natural (LNG):

Es la capacidad de los modelos de inteligencia artificial para producir texto que suena natural y coherente, similar al lenguaje humano. Este proceso implica la creación de oraciones y párrafos a partir de datos o instrucciones, permitiendo que las máquinas se comuniquen efectivamente con los usuarios. La generación de lenguaje natural se utiliza en aplicaciones como chatbots, asistentes virtuales y herramientas de redacción, mejorando la interacción entre humanos y máquinas.

Gobernanza de la IA:

Se refiere al conjunto de políticas, regulaciones y prácticas que guían el desarrollo, implementación y uso responsable de la inteligencia artificial. La gobernanza de la IA busca asegurar que estas tecnologías se utilicen de manera ética, transparente y justa, protegiendo los derechos de los individuos y minimizando riesgos como la discriminación y la invasión de la privacidad. Una gobernanza efectiva es fundamental para fomentar la confianza en la IA y promover su beneficio para la sociedad.

GPT-n (Generative Pre-trained Transformer):

Es una serie de modelos de lenguaje desarrollados por OpenAI que utilizan la arquitectura de transformadores para generar texto de manera coherente y contextual. Estos modelos están preentrenados en grandes volúmenes de texto, lo que les permite entender y producir lenguaje natural. Cada versión (como GPT-3 o GPT-4) mejora en términos de capacidad, precisión y versatilidad, permitiendo aplicaciones en diversos campos como la redacción, la conversación y la asistencia automatizada.

GPT-3:

Es la tercera versión del modelo de lenguaje Generative Pre-trained Transformer desarrollado por OpenAI. Con 175 mil millones de parámetros, GPT-3 es capaz de generar texto coherente y contextual en respuesta a diversas solicitudes. Su entrenamiento en un amplio conjunto de datos le permite realizar tareas como redacción, traducción y generación de diálogos, haciendo de él una herramienta poderosa en aplicaciones de inteligencia artificial, educación y atención al cliente.

GPT-4:

Es la cuarta versión del modelo de lenguaje Generative Pre-trained Transformer de OpenAI, que mejora en comparación con sus predecesores en términos de capacidad, precisión y versatilidad. Con un mayor número de parámetros y un entrenamiento más avanzado, GPT-4 puede generar texto más coherente y contextualizado, entender instrucciones complejas y realizar tareas más sofisticadas. Se utiliza en aplicaciones que requieren interacción más natural y efectiva, como asistentes virtuales, generación de contenido y herramientas de traducción.

H

Heurística:

Se refiere a un enfoque práctico o método de resolución de problemas que utiliza reglas empíricas y juicios aproximados en lugar de técnicas exhaustivas o matemáticas. Las heurísticas ayudan a los algoritmos a encontrar soluciones rápidas a problemas complejos, guiando la búsqueda en espacios de solución grandes. Aunque no siempre garantizan la óptima, las heurísticas son valiosas para mejorar la eficiencia y acelerar el proceso de toma de decisiones en diversas aplicaciones.

Human-on-the-loop (HOTL):

Es un enfoque en el diseño de sistemas de inteligencia artificial donde los humanos participan activamente en el proceso de toma de decisiones, supervisando y ajustando las salidas del modelo. Este método combina la automatización con la intervención humana, asegurando que los resultados sean revisados y validados por personas antes de su implementación. El enfoque HOTL mejora la precisión y la confianza en los sistemas de IA, especialmente en aplicaciones críticas donde la intervención humana puede ser necesaria para garantizar la ética y la seguridad.

I

Identidad digital:

Es la representación en línea de una persona, organización o entidad, que incluye información personal, perfiles en redes sociales, interacciones y datos asociados a su presencia en Internet. Abarca aspectos como la privacidad, la seguridad y la reputación en línea, y es fundamental para la autenticación en servicios digitales.

Gestionar adecuadamente la identidad digital es crucial para proteger la información personal y mantener el control sobre cómo se percibe en el entorno digital.

Inescrutabilidad (Inscrutability):

Se refiere a la dificultad de comprender o interpretar cómo y por qué un modelo de inteligencia artificial toma decisiones específicas. Esta falta de transparencia puede generar desconfianza y preocupación, especialmente en aplicaciones críticas donde las decisiones tienen un impacto significativo. La inescrutabilidad plantea desafíos éticos y prácticos, resaltando la necesidad de desarrollar métodos que hagan los procesos de IA más comprensibles y accesibles para los usuarios.

Inferencia:

Es el proceso mediante el cual un modelo de inteligencia artificial utiliza datos de entrada para hacer predicciones o conclusiones sobre datos no observados. En este contexto, la inferencia implica aplicar el conocimiento adquirido durante el entrenamiento del modelo a nuevos casos, permitiendo que el sistema tome decisiones basadas en la información disponible. La calidad de la inferencia es crucial para el rendimiento del modelo, ya que determina su efectividad en tareas como clasificación, regresión y recomendación.

Ingeniería de prompts:

Es el proceso de diseñar y formular preguntas o instrucciones específicas que se proporcionan a un modelo de lenguaje para obtener respuestas precisas y relevantes. Esta técnica implica experimentar con diferentes formulaciones y estructuras para optimizar las salidas del modelo, mejorando la calidad de la interacción. La ingeniería de prompt es fundamental para maximizar el rendimiento de modelos como ChatGPT, permitiendo a los usuarios aprovechar al máximo sus capacidades de generación de texto.

Inteligencia Artificial (IA):

Es un campo de la informática que se centra en la creación de sistemas y algoritmos capaces de realizar tareas que normalmente requieren inteligencia humana. Estas tareas incluyen el aprendizaje, la razonación, la percepción y la toma de decisiones. La IA utiliza técnicas como el aprendizaje automático y el procesamiento de lenguaje natural para resolver problemas complejos en diversas áreas, como la medicina, la automoción y la atención al cliente, transformando la manera en que interactuamos con la tecnología.

Inteligencia artificial general (IAG):

La inteligencia artificial general (AGI) es un concepto teórico que se refiere a la capacidad

de una máquina de aprender y comprender cualquier tarea intelectual que un ser humano pueda realizar. Se refiere a un sistema de IA que posee una amplia gama de habilidades cognitivas, muy parecidas a las humanas, lo que le permite aprender, razonar, adaptarse a nuevas situaciones y desarrollar soluciones creativas en diversas tareas y dominios, en lugar de estar limitado a tareas específicas como lo están los sistemas de IA estrecha (débil). También es llamada inteligencia artificial fuerte, ya que es una forma teórica que pretende desarrollar capacidades mentales y funciones que imitan al cerebro humano.

Inteligencia Artificial ampliada (Intelligence Augmentation):

La amplificación de la inteligencia o IA aumentada se centra en el desarrollo de tecnología para mejorar las capacidades cognitivas de los seres humanos, sin buscar su reemplazo.

Inteligencia Artificial Agéntica (IAA):

Un modelo de IA diseñado no solo para generar respuestas o análisis pasivos, sino también para tomar decisiones, actuar de manera autónoma y ejecutar tareas de forma proactiva.

Inteligencia Artificial débil (Artificial Narrow Intelligence/ ANI):

La inteligencia artificial débil o estrecha es aquella que aplica técnicas y algoritmos de aprendizaje automático y procesamiento del lenguaje natural para realizar tareas muy específicas de forma automatizada, por lo que tienen capacidad limitada fuera de ese rango de tareas. Se centra en tareas definidas, tales como la identificación de patrones, reconocimiento de imágenes, etcétera. Se puede encontrar en diversas aplicaciones como chatbots, vehículos autónomos, entre otros.

Inteligencia Artificial ética:

Se encarga de estudiar y dar respuesta a los dilemas éticos y sociales relacionados con el diseño, desarrollo e implementación de la IA.

Inteligencia Artificial conversacional:

Se centra en elaborar un conjunto de programas que trabajan juntos que puedan entender y generar un lenguaje similar al humano y llevar a cabo una conversación de ida y vuelta. Un ejemplo de ello sería un chatbot que pudiera entender y responder a consultas de clientes de manera similar a una conversación entre personas.

Inteligencia Artificial de Código Abierto (Stable Diffusion):

Utiliza aprendizaje profundo para generar imágenes a partir de indicaciones de texto natural.

Inteligencia Artificial Generativa (IAGen):

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.

Integración:

Se dice que hay una integración cuando un programa/producto armoniza con las capacidades y beneficios de la IA para mejorar aspectos diversos (funcionalidad, rendimiento, personalización, seguridad, etc.). Por lo tanto, no debe verse como dos herramientas por separado, sino como una incorporación de ambos.

Iteración:

Proceso continuo de retroalimentación entre la persona y la IA que permite la personalización del contenido, la detección y corrección de errores, la exploración y el trabajo creativo conjunto para generar contenidos, ideas o soluciones.

L

Lógica Difusa:

Método basado en grados de verdad en lugar de valores binarios (verdadero/falso), útil en IA para modelar incertidumbre.

Lenguaje digital:

Conjunto de formas de comunicación y expresión que se generan y procesan mediante tecnologías digitales, especialmente a través de la inteligencia artificial (IA). Incluye texto, imágenes, videos y audio que pueden ser creados, analizados y transformados por algoritmos y modelos de IA. Este lenguaje abarca desde la generación automática de contenido hasta interacciones en tiempo real, permitiendo una comunicación más eficiente, personalizada y accesible en el entorno digital.

M

Marco legal de la IA:

Conjunto de leyes y normativas que regulan el desarrollo y uso de la IA, incluyendo la protección de datos personales, la responsabilidad por daños causados por la IA y la propiedad intelectual de los algoritmos.

Memoria asociativa:

Es la capacidad de un sistema para almacenar, recuperar y procesar información relacionada basada en conexiones entre elementos, lo que le permite identificar y utilizar de manera eficiente los datos relevantes para la toma de decisiones.

Metadatos/metadata:

Los metadatos son datos contenidos dentro de archivos; son datos utilizados para definir, contextualizar o caracterizar a aquellos archivos; describiendo sus datos. Se pueden diferenciar en tres categorías: metadatos que sirven para descubrir e identificar el contenido de un dato, metadatos que explican cómo se organizan los distintos

componentes de un dato y . metadatos que aluden a la fuente de la información.

Algunos de los tipos de información que proporcionan los metadatos son información de aspectos formales, información de copyright o derechos de autor, información del contenido, información del contexto y datos sobre la autenticidad del recurso o documento.

Minimización de Riesgos Algorítmicos:

Estrategias para reducir sesgos y errores en modelos de IA.

Modelo base:

Amplia categoría de modelos de IA que incluyen modelos de lenguaje grandes y otros tipos de modelos como los de visión por computadora y modelos de aprendizaje por refuerzo. Se les llama modelos “base” porque sirven como el pilar sobre el que construir aplicaciones, atendiendo a una amplia gama de dominios y casos de uso.

Modelo determinista:

Un modelo determinista sigue un conjunto específico de reglas y condiciones para alcanzar un resultado definitivo, operando en una base de causa y efecto.

Modelo discriminativo:

Los modelos discriminativos son algoritmos diseñados para modelar y aprender directamente el límite entre diferentes clases o categorías en un conjunto de datos.

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. Los modelos de lenguaje grande son de "caja negra" por naturaleza.

Modelo de IA:

Representación matemática entrenada en datos para realizar predicciones o tomar decisiones.

Modelo de aprendizaje automático:

Inspirado en la estructura y función del cerebro humano que está compuesto por capas de nodos interconectados o “neuronas”. Permiten reconocer patrones y resolver problemas, basados en formación automática o aprendizaje profundo, buscando modelar los mecanismos de análisis del cerebro humano.

Modelo de Lenguaje de gran tamaño (o grande):

Sus siglas en inglés son “LLM- Large Language Model”. Son sistemas de IA generativa que pueden procesar lenguaje natural bajo la forma de una conversación. Han sido entrenados con enormes cantidades de datos, por lo que logran capturar las estructuras,

patrones y relaciones presentes en el lenguaje humano para poder generar texto coherente y relevante bajo una lógica probabilística, prediciendo la próxima mejor palabra o conjunto de palabras en una oración a partir de una instrucción dada por un usuario en lenguaje natural, considerando el contexto de ese pedido. Hay modelos muy conocidos, como BERT, PaLM, GPT-2, GPT-3, GPT-3.5 y el innovador GPT-4. Todos estos modelos varían en tamaño (número de parámetros que se pueden ajustar), en la amplitud de tareas (codificación, chat, científicas, etc.), y en el conocimiento en que han sido entrenados.

Modelado de secuencias:

Un subcampo del procesamiento de lenguaje natural que se enfoca en modelar datos secuenciales como texto, habla o datos de series temporales. En este caso, un modelo de secuencia sería aquel que puede predecir la próxima palabra en una oración o generar texto coherente.

N

Neuronas Artificiales:

Unidades básicas de las redes neuronales inspiradas en las neuronas biológicas.

Normalización:

Técnica para escalar datos y mejorar la eficiencia de modelos de IA.

O

Ofuscación (Obfuscation):

Distintas formas en las que la IA, sus procesos y/o sus resultados resultan oscuros para sus creadores, operadores y/o destinatarios.

One-shot prompting:

La solicitud busca la creación de un contenido específico, como un texto o una imagen. Por ello, se proporciona un ejemplo del resultado deseado. Por ejemplo: "Redactá una cláusula de rescisión anticipada de un contrato de locación, similar a esta cláusula que regula la finalización anticipada en el contrato que adjunto".

Opacidad (Opacity):

Falta de transparencia en el acceso a comprender su funcionamiento sea por razones de complejidad técnica para quienes carecen de conocimientos en el área (opacidad técnica) o por dispositivos legales que lo transformen en una caja negra (opacidad legal).

OpenAI :

Es la empresa que desarrolló ChatGPT. Tiene como objetivo desarrollar y promover de manera responsable una IA amigable.

Optimización:

Es el proceso de ajustar los parámetros de un modelo para minimizar una función de pérdida que mide la diferencia entre las predicciones del modelo y los valores verdaderos, buscando mejorar su rendimiento.

P

Parámetros:

En el aprendizaje automático, son las variables internas que el modelo utiliza para hacer predicciones. Se aprenden de los datos de entrenamiento durante el proceso de entrenamiento. Como ejemplo, en una red neuronal, los pesos y los sesgos son parámetros.

Perceptrón:

Modelo matemático de neurona artificial utilizado en aprendizaje automático.

Perfilado:

Se trata del procesamiento de datos personales con el fin de evaluar determinados aspectos de la vida de una persona física (situación económica, salud, preferencias personales, etc.).

Pervasividad (Pervasivity):

Imposibilidad de decidir sobre ser o no sujeto de aplicación de sistemas de IA o de escapar a su uso o a las consecuencias de su uso.

Plugin de IA:

Es un componente de software especializado que permite a los sistemas de IA interactuar con aplicaciones y servicios externos.

Predicción:

Proceso mediante el cual un modelo de IA estima un valor futuro basándose en datos previos.

Preentrenamiento:

La fase inicial de entrenamiento de un modelo de aprendizaje automático donde el modelo aprende características generales, patrones y representaciones de los datos sin conocimiento específico de la tarea a la que se aplicará más tarde. Este proceso permite que el modelo desarrolle una comprensión fundamental de la distribución subyacente de los datos y extraiga características significativas que pueden ser aprovechadas para el ajuste fino posterior en tareas específicas.

Prejuicio:

Ver sesgo algorítmico.

Privacidad por Diseño (Privacy by Design):

Enfoque que integra la protección de datos personales desde las etapas iniciales de diseño y desarrollo de sistemas y aplicaciones, garantizando la privacidad como un principio fundamental.

Portabilidad:

Es el derecho del afectado a recibir los datos personales que le incumban, que haya facilitado a un responsable del tratamiento, en un formato estructurado, de uso común y lectura mecánica, y a transmitirlos a otro responsable del tratamiento sin que lo impida el responsable al que se los hubiera facilitado.

Procesamiento del Lenguaje Natural (NLP):

El procesamiento del lenguaje natural (PLN) es un subcampo de la informática y la inteligencia artificial (IA) que utiliza el machine learning para permitir que los ordenadores entiendan y se comuniquen con el lenguaje humano.

Procesamiento de voz:

El procesamiento de voz en IA se refiere a la secuencia de conversión de voz a texto seguida de la síntesis de texto a voz.

Prompt:

Es una instrucción o conjunto de instrucciones (input) proporcionada en lenguaje natural que un usuario le da a un modelo de IA generativa para que genere una respuesta. Puede incluir información contextual específica, componentes esenciales y basarse en técnicas de prompting que guían la generación del contenido deseado para lograr resultados más efectivos y eficientes.

Prompts recurrentes:

Son una estrategia para guiar modelos de IA y así producir resultados de mayor calidad. Implica proporcionar al modelo una serie de prompts o preguntas que se basan en respuestas anteriores, refinando tanto el contexto como la comprensión de la IA para lograr el resultado deseado.

Prompteo:

Instrucción o frase inicial que se le proporciona al modelo de lenguaje para generar un texto y obtener un resultado específico en distintos formatos.

R**Razonamiento:**

El razonamiento de la IA es el proceso por el cual los sistemas de inteligencia artificial resuelven problemas, piensan críticamente y crean nuevo conocimiento analizando y procesando la información disponible, lo que les permite tomar decisiones bien

informadas en diversas tareas y dominios.

Razonamiento multi-salto :

Es un término que se utiliza a menudo en el procesamiento del lenguaje natural, y más específicamente en aquellas tareas relacionadas con la comprensión de lectura de máquinas. Se refiere al proceso por el cual un modelo de IA recupera respuestas a preguntas conectando múltiples piezas de información presentes en un texto dado o a través de varias fuentes y sistemas, en lugar de extraer directamente la información de un solo pasaje.

Reconocimiento automático de voz (ASR):

Es una tecnología que transcribe el lenguaje hablado a texto.

Reconocimiento de Imágenes:

Capacidad de un sistema de IA para identificar objetos, personas o patrones en imágenes.

Redes generativas adversarias/adversariales (RGAs/ GANs):

Las redes generativas adversarias o antagónicas son un ejemplo poderoso de red neuronal, capaz de generar información nueva nunca vista que se asemeja mucho a los datos de entrenamiento.

Red neuronal:

Se trata de sistemas compuestos de un gran número de elementos procesadores, fuertemente interconectados entre sí, que procesan información mediante una respuesta dinámica a las entradas exteriores recibidas. Posee dos elementos constituyentes esenciales, los procesadores y sus interconexiones.

Los procesadores, que equivalen a las neuronas cerebrales, son dispositivos que reciben un número de señales de entrada o inputs y pueden generar o no una señal de salida u outputs. En el caso de que se genere una señal de salida, esta es remitida a otros elementos procesadores como nueva señal de entrada, usando para ello las interconexiones existentes entre ellos.

Los elementos procesadores se disponen en arreglos lineales llamados capas. Las capas pueden ser de distintas clases: de entrada, de salida y ocultas. Las capas de entrada reciben señales desde el exterior, las capas de salida envían señales hacia el exterior, y las ocultas, no interactúan con el medio exterior, y sus entradas y salidas se producen dentro del sistema. La interconexión de los elementos procesadores es tal que estas conexiones constituyen un patrón de conectividad. Este patrón es el que permite que el sistema se entrene para un aprendizaje dado y a su vez también determina la respuesta a cualquier entrada arbitraria.

Red Neuronal Convolucional (CNN):

Arquitectura de redes neuronales utilizada en procesamiento de imágenes y visión por computadora.

Red Neuronal Recurrente (RNN):

Tipo de red neuronal usada en procesamiento de secuencias, como el análisis de texto o series temporales.

Regresión:

Técnica estadística utilizada en IA para modelar relaciones entre variables.

Responsabilidad algorítmica:

Principio que establece la responsabilidad legal y ética de los desarrolladores, proveedores y usuarios de sistemas de IA por los daños o perjuicios causados por sus algoritmos.

Responsabilidad demostrada o proactiva (accountability):

Principio que indica que se debe poder demostrar la licitud del tratamiento del dato personal.

Retropagación:

Es un algoritmo frecuentemente utilizado en el entrenamiento de redes neuronales artificiales. Ajusta las conexiones entre neuronas para minimizar el error en la predicción de un modelo. Es efectivo para, por ejemplo, el reconocimiento de patrones o la clasificación de imágenes.

S**Sesgo algorítmico (bias):**

Refiere a los prejuicios, estereotipos, relaciones de poder y/o valores sociales negativos, entre otros, presentes en los resultados generados por un sistema de IA generativa, que se originan en los datos de entrenamiento, estructuras del modelo o decisiones de diseño. Afecta los resultados de un modelo de inteligencia artificial, a menudo debido a prejuicios en los datos de entrenamiento. Este sesgo puede llevar a decisiones injustas o erróneas, reflejando desigualdades existentes en la sociedad.

Seudonimización:

El tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identifiable.

Singularidad:

También llamada singularidad tecnológica, se refiere a un punto hipotético futuro en el tiempo cuando el crecimiento tecnológico se vuelve incontrolable e irreversible, llevando a cambios imprevisibles en la civilización humana.

Sistema inteligente de tutoría (Intelligent Tutoring Systems ITS):

Es un programa de IA que simula ser un tutor/profesor que brinda una experiencia personalizada. Puede ser usado para varios escenarios en educación: monitoreo, asesoría, retroalimentación y simulaciones de entornos de aprendizaje.

Sobreajuste:

Un problema que ocurre cuando un modelo es demasiado complejo, rindiendo bien en los datos de entrenamiento, pero mal en datos no vistos.

T

Texto a voz (TTS):

Es una tecnología que convierte texto escrito en salida de voz hablada. Permite a los usuarios escuchar contenido escrito siendo leído en voz alta, típicamente utilizando voz sintetizada.

Transformador (transformer):

Representa un tipo de arquitectura de red neuronal diseñada para procesar datos secuenciales, como es el texto. La arquitectura transformadora aprende contexto y en consecuencia significado. Se utiliza en modelos avanzados de procesamiento del lenguaje natural como GPT o BERT.

Tratamiento de Datos Personales:

Cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción

Tree of Thoughts:

El objetivo del prompt es obtener conclusiones bien fundamentadas o desarrollar un pensamiento crítico sobre un tema específico que requiera un análisis detallado. Esta técnica se denomina así por su estructura jerárquica y ramificada, y busca representar ideas y argumentos de forma organizada. Por ejemplo para un determinado caso de custodia de un menor: “Desarrollá un árbol de pensamiento que explore los argumentos a favor de otorgar la custodia principal a la madre, considerando factores como el vínculo afectivo primario y la disponibilidad horaria. Luego, explorá los argumentos a favor del padre, considerando su estabilidad económica y la cercanía a la escuela del menor. Finalmente, considera un punto de vista que proponga la custodia compartida, analizando los beneficios y desafíos de esta modalidad para el bienestar del menor”.

V

Validación cruzada :

Es un método utilizado para evaluar los resultados de un análisis y garantizar la exactitud de las predicciones de los modelos de aprendizaje automático.

Visión computacional (computer vision):

Es un campo de la IA que le permite a los sistemas analizar el contenido visual (imágenes o videos), con el objetivo de que puedan tomar decisiones o hacer recomendaciones basadas en los contenidos analizados.

Z

Zero-shot prompting:

Utilizado en situaciones cuando se requieren respuestas a preguntas básicas o sobre temas amplios. El objetivo es proporcionar información de manera concisa y sin una introducción contextual detallada. Por ejemplo: “¿Cuál es la capital de la Provincia de Buenos Aires?”.

BIBLIOGRAFIA

1. [Léxico de la inteligencia artificial](#) de la UNESCO.
2. [Glosario de inteligencia artificial \(IA\) para la educación](#) del Observatorio del Instituto para el futuro de la educación, Tecnológico de Monterrey.
3. [Glosario sobre IA: conceptos básicos](#) del eLearning Innovation Center (eLinC), Universidad Abierta de Cataluña.
4. [Glosario y guía básica de inteligencia artificial](#) de la Prosecretaría de Políticas Digitales de la Facultad de Ciencias Jurídicas y Sociales, Universidad Nacional de La Plata.
5. [Glosario de términos de inteligencia artificial \(IA\) y aprendizaje automático \(AA\)](#) de Chubut Digital.
6. [Glosario de inteligencia artificial](#) del Proyecto IDIS, Facultad de Arquitectura Diseño y Urbanismo, Universidad de Buenos Aires.
7. [Glosario de Inteligencia Artificial \(IA\) para la educación](#), del Instituto para el Futuro de la Educación, Tecnológico de Monterrey, Nuevo León, México.
8. <https://espanadigital.gob.es/lineas-de-actuacion/programa-nacional-de-algoritmos-verdes>
9. <https://www.coe.int/en/web/artificial-intelligence/glossary>

Descargo de responsabilidad

Las opiniones expresadas en esta publicación son responsabilidad exclusiva de sus autores y autoras, y no reflejan necesariamente la posición institucional del Colegio Público de la Abogacía de la Capital Federal. El CPACF no se responsabiliza por la exactitud o integridad del contenido, ni por el uso que terceros hagan de él.

Créditos / Editorial

Documento elaborado por el Instituto de Derecho Informático del CPACF.

Coordinación: Dr. Eduardo Molina Quiroga.

Redacción: Dres. Eduardo Molina Quiroga, Clara Costamagna, Oscar Vidal, Gonzalo Carrasco, Tatiana Fij, Mariela Aciar, Leonor Guini, Horacio Granero, Cecilia Cascante y Claudio Schifer.

Buenos Aires, julio de 2025