Inteligencia Artificial y el Estado de Derecho en la Administración de Justicia

Ricardo Scarpa (derechoartificial.com)

Capítulo 1: Arquitectura Técnica y Componentes Básicos de los Sistemas de IA

Este capítulo establece la base técnica necesaria para comprender la tecnología antes de abordar su regulación.

- **1.1. Definiciones fundamentales y evolución sociotécnica:** Concepto de IA según la UNESCO y otros organismos internacionales.
- **1.2. El ciclo de vida del aprendizaje automático (Machine Learning):** Desde la definición de objetivos hasta la integración del modelo.
- **1.3. La importancia crítica de la infraestructura de datos:** Datificación, etiquetado y el principio FAIR (Localizable, Accesible, Interoperable y Reutilizable).
- **1.4. Seguridad sistémica:** Gestión de riesgos de ciberseguridad y protección contra el envenenamiento de datos.

Capítulo 2: Ecosistema de Aplicaciones de la IA en el Poder Judicial

Se analiza cómo la técnica se convierte en herramienta operativa dentro del sistema de justicia.

- **2.1. Herramientas de soporte administrativo y procesal:** Descubrimiento electrónico (e-discovery), revisión de documentos y gestión de archivos digitales.
- **2.2. Procesamiento de Lenguaje Natural (PLN)**: Aplicaciones en traducción de sentencias y transcripción en vivo de audiencias.
- **2.3. IA Generativa en el ámbito legal**: Uso de modelos de lenguaje de gran tamaño (LLM) para la redacción de borradores y desafíos de "alucinación".
- **2.4. Análisis predictivo y Toma de Decisiones Algorítmicas (ADM)**: Evaluación de riesgos y asistencia en la resolución de conflictos.

Capítulo 3: Análisis de Implementación Global: Estudios de Caso

Transición de la teoría a la práctica mediante el examen de despliegues reales.

- **3.1. Automatización en tribunales superiores:** El caso del sistema VICTOR en Brasil para el análisis de apelaciones.
- **3.2. Sistemas de redacción y apoyo jurisdiccional:** La experiencia de Prometea en Argentina y PretorIA en Colombia.

- **3.3. IA para la eficiencia y el acceso:** El portal SUPACE en la India y el bot jurado de Los Ángeles.
- **3.4. Herramientas de evaluación de riesgos penales:** El sistema HART en el Reino Unido y el debate sobre el algoritmo COMPAS en EE. UU..

Capítulo 4: Ética Algorítmica y Desafíos de la Opacidad (La "Caja Negra")

Este capítulo actúa como puente entre la implementación técnica y el marco de derechos humanos.

- **4.1. Marco ético de la UNESCO:** Reflexión normativa y principios fundamentales para una IA confiable.
- **4.2. El problema de la "Caja Negra" e IA Explicable (XAI):** La lucha contra la opacidad de los algoritmos de autoaprendizaje.
- **4.3. Sesgos algorítmicos y discriminación:** Sesgos de muestra, de asociación y de automatización.
- **4.4. Tecnologías de alto riesgo:** Identificación biométrica, reconocimiento facial y el ecosistema de las falsedades profundas (deepfakes).

Capítulo 5: Impacto de la IA en los Derechos Humanos y Garantías Procesales

Análisis regulatorio profundo sobre la protección del individuo frente al sistema.

- **5.1. El debido proceso y el derecho a un juicio justo:** Desafíos de la presunción de inocencia ante la vigilancia predictiva.
- **5.2. Derecho a la privacidad y protección de datos:** El impacto de la elaboración de perfiles y el seguimiento en línea.
- **5.3. Libertad de expresión y acceso a la información:** Moderación de contenidos, desinformación y el efecto "escalofriante" de la vigilancia.
- **5.4. Derecho a un recurso efectivo:** Mecanismos de impugnación de decisiones automatizadas y el "derecho a la explicación".

Capítulo 6: Modelos de Gobernanza, Regulación y el Futuro del Estado de Derecho

Conclusión multidisciplinar que propone marcos de control y supervisión.

- **6.1. Enfoques de gobernanza basados en el riesgo:** El ejemplo del proyecto de Ley de IA de la Unión Europea y la clasificación de niveles de riesgo.
- **6.2. El principio del "Humano en el Circuito" (HITL):** La necesidad de supervisión y validación humana efectiva en las decisiones judiciales.
- **6.3. Evaluaciones de impacto:** Metodologías HRIA (Impacto en Derechos Humanos) y FRAIA (Derechos Fundamentales y Algoritmos).
- **6.4. El rol del operador judicial:** El juez como garante último del Estado de derecho frente a la expansión tecnológica.

Capítulo 1: Arquitectura Técnica y Componentes Básicos de los Sistemas de IA

Cuando nos adentramos en el estudio de la Inteligencia Artificial (IA) aplicada al derecho, es fácil caer en la tentación de verla como una especie de oráculo tecnológico, una entidad casi mística. Sin embargo, para nosotros, como operadores del sistema de justicia, es imperativo despojarla de ese halo de misterio y entenderla por lo que es: un **sistema sociotécnico**. No hablamos simplemente de máquinas, sino de un entramado donde la capacidad de procesar datos para asemejarse al comportamiento inteligente —el razonamiento, el aprendizaje o la predicción— se entrelaza con decisiones humanas previas que configuran su funcionamiento.

1.1. Definiciones fundamentales y evolución sociotécnica

Debemos partir de una base clara. La IA no es un concepto estático; de hecho, la propia UNESCO reconoce que cualquier definición debe ser lo suficientemente flexible para evolucionar con los avances tecnológicos. En esencia, estamos ante tecnologías de procesamiento de información que integran modelos y algoritmos para realizar tareas cognitivas que antes eran exclusivas de los seres humanos. Es curioso observar cómo instituciones como la OCDE o la ISO ponen el foco en la capacidad de estos sistemas para influir en entornos reales o virtuales mediante predicciones o recomendaciones, mientras que otros organismos enfatizan su autonomía operativa.

A menudo cometemos el error de usar "IA" y "Aprendizaje Automático" (o *Machine Learning*) como sinónimos. Me gustaría puntualizar que la IA es un campo mucho más vasto. Si bien el aprendizaje automático es el motor que está impulsando la revolución actual, la IA también abarca áreas como la planificación y el razonamiento simbólico. No es solo cuestión de procesar números; es el intento de modelar el conocimiento mismo.

Si miramos atrás, la evolución de esta tecnología se divide en lo que llamamos "olas". La primera consistió en **sistemas expertos**, basados en reglas rígidas del tipo "si sucede A, entonces haz B" (árboles de decisión binarios, por ejemplo). Son útiles, sí, pero incapaces de lidiar con lo inesperado. La segunda ola, donde nos encontramos de lleno, es la del aprendizaje automático: aquí el sistema infiere sus propias reglas a partir de los datos. El horizonte es la tercera ola, la de la **adaptación contextual**, donde la máquina no solo decidirá, sino que podrá explicarnos por qué lo hizo, superando finalmente el muro de los **algoritmos opacos**.

1.2. El ciclo de vida del aprendizaje automático (Machine Learning)

El aprendizaje automático es, en esencia, un conjunto de técnicas que permite a las máquinas aprender mediante patrones y deducciones. Para que un modelo sea útil en un juzgado —pensemos en una herramienta de transcripción en vivo o de análisis de precedentes— debe atravesar un ciclo de vida riguroso. Este proceso comienza

con una definición clara de objetivos (¿qué problema queremos resolver realmente?) y sigue con la recopilación y preparación de datos.

Me parece fundamental recalcar la fase de **etiquetado de datos**. Es el proceso de dar contexto a la información bruta: decirle a la máquina que esta imagen es un "tumor" o que este texto es un "contrato". Sin este entrenamiento —especialmente en el aprendizaje supervisado— el modelo es incapaz de diferenciar entre lo relevante y lo accesorio. Una vez entrenado, el modelo debe probarse con datos que no ha visto antes para verificar su precisión. Solo entonces se integra en el flujo de trabajo judicial. Pero ojo, la eficacia aquí depende de tres variables: la cantidad de datos, su calidad y la potencia de cálculo. Si alguna falla, el sistema entero se tambalea.

1.3. La importancia crítica de la infraestructura de datos

Llegamos a lo que considero el corazón del sistema: los datos. Vivimos en plena **datificación**, un proceso donde cada movimiento, cada "me gusta" y cada sentencia se convierte en una huella digital que los algoritmos pueden evaluar. Pero no cualquier dato sirve. Para que un sistema judicial sea justo, su infraestructura debe seguir el **principio FAIR**: la información debe ser Localizable, Accesible, Interoperable y Reutilizable.

El problema —y esto es algo que me preocupa profundamente— es que si alimentamos al sistema con datos de baja calidad, inexactos o incompletos, el resultado será inevitablemente sesgado. Muchos sistemas de justicia penal utilizan modelos estadísticos basados en registros policiales que ya arrastran desigualdades sociales históricas. La IA no borra el sesgo; a menudo lo amplifica. Por eso, antes de certificar cualquier modelo para el mercado judicial, debemos exigir que los datos se hayan recopilado de forma ética y que representen fielmente a la población afectada. La **"invisibilidad de los datos"** de grupos marginados es una receta directa para la discriminación automatizada.

1.4. Seguridad sistémica: un requisito de integridad

Finalmente, no podemos hablar de arquitectura técnica sin abordar la **seguridad sistémica**. La ciberseguridad aquí no es solo proteger una contraseña; es gestionar riesgos que afectan a la integridad de las decisiones judiciales. Los sistemas de IA tienen vulnerabilidades únicas. Una de las más insidiosas es el **envenenamiento de datos** durante la etapa de entrenamiento, donde un atacante manipula el conjunto de datos para que el sistema "aprenda" patrones defectuosos de forma intencionada.

También existen los ataques de entrada mediante **ejemplos antagónicos**. Son perturbaciones tan sutiles que el ojo humano no las detecta, pero que confunden al algoritmo (como un simple adhesivo en una señal de tráfico que hace que un coche autónomo la ignore). Si estas técnicas se aplicaran para engañar a un sistema de reconocimiento de imágenes médicas o de vigilancia, las consecuencias serían,

literalmente, fatales. Por tanto, la regulación debe ser prescriptiva: las instituciones deben estar obligadas a proteger sus sistemas contra virus, accesos no autorizados y manipulaciones de modelos en cada fase, desde la preparación del dato hasta la implementación final. Al final del día, la tecnología solo es una herramienta útil si podemos confiar en que nadie ha alterado los cimientos sobre los que se apoya.

Capítulo 2: Ecosistema de Aplicaciones de la IA en el Poder Judicial

Habiendo desgranado la arquitectura técnica en el capítulo anterior, es momento de observar cómo este entramado de datos y algoritmos aterriza, a veces con más ruido que nueces, en la práctica judicial diaria. No estamos ante una promesa futurista; la IA ya está operando en nuestros tribunales bajo diversas formas, desde la gestión documental más rutinaria hasta la compleja, y a menudo polémica, evaluación de riesgos penales. Sin embargo, como operadores del sistema, debemos evitar la fascinación tecnológica y analizar estas herramientas bajo el prisma de la eficiencia, pero también de la integridad procesal.

2.1. Herramientas de soporte administrativo y procesal: El fin de los expedientes interminables

Uno de los mayores cuellos de botella de cualquier sistema de justicia es la ingente cantidad de **Información Almacenada Electrónicamente (IAE)** que se debe procesar en cada litigio. Aquí es donde el **descubrimiento electrónico (e-discovery)** y la revisión de documentos han marcado un antes y un después. No se trata simplemente de buscar palabras clave —lo cual ya hacíamos con herramientas básicas—, sino de utilizar técnicas de aprendizaje automático no supervisado, como la agrupación o *clustering*, para identificar patrones y conceptos similares.

Me resulta interesante ver cómo la **Revisión Asistida por Tecnología (TAR)** permite que, tras una codificación inicial realizada por humanos, la máquina aprenda a distinguir lo relevante de lo superfluo en volúmenes de datos que a una persona le tomaría años procesar. Esto, que aumenta la velocidad de revisión entre un 15 % y un 20 %, no solo es una cuestión de ahorro de costes, sino de acceso a la justicia: un proceso más rápido es, por definición, un proceso más justo. No obstante, surge una duda razonable que los tribunales ya están empezando a debatir: ¿cuáles serán los estándares de admisibilidad para estas ideas generadas por IA que los humanos decidimos aceptar o rechazar?

2.2. Procesamiento de Lenguaje Natural (PLN): Derribando barreras lingüísticas

El **Procesamiento de Lenguaje Natural (PLN)** es, quizás, la herramienta que más directamente impacta en la transparencia del sistema. Pensemos en el caso de la India con el software **SUVAS**, diseñado para traducir sentencias a nueve idiomas locales. Esta aplicación de la IA no es un lujo técnico; es una necesidad democrática para asegurar que el ciudadano entienda las órdenes que afectan su vida.

Del mismo modo, la transcripción en vivo de audiencias está ganando terreno. Sin embargo, aquí debemos ser sumamente cautelosos. Los modelos de PLN, aunque potentes, tienen una reputación —bien ganada, por cierto— de fallar ante acentos específicos o giros lingüísticos locales. Un error de traducción en una diligencia judicial no es un simple malentendido; puede ser una vulneración directa de los **derechos fundamentales** del procesado si esa transcripción defectuosa acaba cimentando una resolución.

2.3. IA Generativa en el ámbito legal: Entre la eficiencia y la alucinación

No podemos ignorar la "fiebre" reciente por la **IA generativa** y los modelos de lenguaje de gran tamaño (**LLM**), como ChatGPT. Su capacidad para redactar borradores, resumir jurisprudencia o incluso proponer argumentos legales es asombrosa, pero peligrosamente seductora. Ya hemos visto casos en Colombia, Perú y México donde jueces han recurrido a estas herramientas para motivar decisiones o ilustrar argumentos.

El riesgo aquí es doble. Por un lado, está el fenómeno de las **alucinaciones**, donde el sistema inventa citas o precedentes con una seguridad pasmosa (el caso *Mata vs. Avianca* en EE. UU. es el recordatorio perfecto de este peligro). Por otro lado, existe la preocupación de que los jueces se conviertan en usuarios pasivos, delegando su capacidad de pensamiento crítico en un código que, al final del día, no tiene lealtad al **Estado de derecho** ni ha prestado juramento alguno. La prescripción es clara: cualquier texto generado por IA debe pasar por una certificación y verificación humana exhaustiva utilizando fuentes tradicionales antes de ser presentado ante un tribunal.

2.4. Análisis predictivo y Toma de Decisiones Algorítmicas (ADM): El reto de la objetividad

Finalmente, llegamos al terreno más pantanoso: el **análisis predictivo** y el soporte a la **Toma de Decisiones Algorítmicas (ADM)**. Se nos dice que estas herramientas, como **COMPAS** o **HART**, pueden eliminar el sesgo humano en las decisiones sobre fianzas o sentencias al proporcionar una puntuación de riesgo "objetiva".

Pero cuidado. Es un error de bulto pensar que un algoritmo es neutral por el simple hecho de ser matemático. Como ya hemos apuntado, si el modelo se entrena con datos históricos que arrastran sesgos raciales o socioeconómicos, la IA no corregirá la injusticia; la automatizará y, lo que es peor, la ocultará tras una capa de **algoritmos opacos**. En el sistema de justicia penal, una predicción estadística nunca puede ser la única base para un arresto o una condena. La presunción de inocencia exige que cada caso se analice por sus propios méritos, manteniendo siempre a un **humano en el circuito** que sea capaz de evaluar críticamente —y contradecir, si es necesario— el resultado del sistema. El juez, en definitiva, no debe ser un simple aplicador de algoritmos, sino su evaluador último y garante de los derechos ciudadanos.

Capítulo 3: Análisis de Implementación Global: Estudios de Caso

Tras haber analizado el potencial teórico y el ecosistema de aplicaciones de la Inteligencia Artificial, es momento de poner los pies en la tierra. No estamos hablando de castillos en el aire; la IA ya está dictando el ritmo en despachos y estrados de medio mundo. Sin embargo, al observar estos despliegues, uno se da cuenta de que la brecha entre la eficiencia administrativa y el respeto a los **derechos fundamentales** es, a veces, peligrosamente estrecha. Me propongo analizar en este capítulo cómo diversos Estados han integrado estos sistemas, pasando de la fascinación inicial a una necesaria cautela regulatoria.

3.1. Automatización en tribunales superiores: El caso VICTOR en Brasil

Empecemos por el Tribunal Supremo de Brasil (STF), que se enfrentó a un dilema casi inmanejable: procesar más de cincuenta mil apelaciones anuales. La solución fue **VICTOR**, un sistema diseñado para automatizar el examen de "repercusión general". Lo que antes le tomaba cuarenta minutos a un funcionario —leer y clasificar precedentes vinculantes—, este sistema lo resuelve en segundos.

Lo que me parece técnicamente reseñable es que VICTOR no solo clasifica; ha tenido que aprender a lidiar con una "selva" de formatos, desde PDFs no estructurados hasta documentos no indexados provenientes de toda la geografía brasileña. Es un ejemplo masivo de cómo la IA puede desatascar la maquinaria judicial, pero no nos engañemos: la eficacia aquí depende totalmente de una base de datos de casi tres millones de expedientes. La pregunta que queda en el aire es si esa velocidad no termina por invisibilizar matices que un ojo humano, aunque más lento, detectaría en una apelación compleja.

3.2. Sistemas de redacción y apoyo jurisdiccional: Prometea y PretorIA

En nuestra región, el caso de **Prometea**, en la Fiscalía de Buenos Aires, ha sido la punta de lanza. Su capacidad para reducir un proceso de preparación de juicio de 167 días a tan solo 38 es, sencillamente, impresionante. Se presenta como un sistema experto multifuncional que permite a los jueces, mediante un chatbot o comandos de voz, generar borradores de opiniones legales tras responder apenas cinco preguntas.

No obstante, y aquí es donde la transición a lo regulatorio se vuelve crítica, el despliegue de Prometea en Colombia para la Corte Constitucional (bajo el nombre de **PretorIA**) encendió todas las alarmas sobre la **opacidad**. Hubo una preocupación legítima por el manejo de datos sensibles de víctimas menores de edad o delitos sexuales. Resulta fascinante observar cómo, ante las críticas por la naturaleza de "caja negra" de las redes neuronales, la Corte decidió pivotar hacia una tecnología de "modelado de temas". Esto es una lección de humildad tecnológica: a veces, un modelo menos "inteligente" pero más **explicable** y rastreable es el único compatible con el **Estado de derecho**.

3.3. IA para la eficiencia y el acceso: De la India a Los Ángeles

Si miramos hacia Asia, el sistema **SUPACE** en la India nos muestra un enfoque diferente. Aquí, la IA no decide; se limita a catalogar y procesar la matriz de hechos para que el juez pueda realizar una investigación dinámica de precedentes. Es un asistente, no un sustituto. Junto a él, el software **SUVAS** cumple una función que considero esencialmente democrática: traducir sentencias a nueve idiomas locales para asegurar que el ciudadano entienda lo que se ha fallado sobre su vida.

Por otro lado, me llama la atención el **bot jurado** del Tribunal Superior de Los Ángeles. En una ciudad donde la gente esperaba horas por una multa de tráfico, este asistente procesa ahora 5.000 ciudadanos por semana en cinco idiomas. Es la IA al servicio de la micro-justicia. Son herramientas que, aunque parecen menores, sostienen la confianza pública en el sistema al eliminar la frustración burocrática.

3.4. Herramientas de evaluación de riesgos penales: HART y el dilema de la libertad

Llegamos al terreno más espinoso: la evaluación de riesgos para decidir libertades. En el Reino Unido, la policía de Durham utiliza **HART**, un algoritmo que clasifica a los sospechosos según su probabilidad de reincidencia. Pero cuidado, porque aquí la técnica choca frontalmente con la ética procesal. Se ha señalado que HART es propenso a "criminalizar en exceso" porque está programado para subestimar a quienes califican para programas de rehabilitación. Esto, a mi juicio, es una afrenta directa al principio *in dubio reo*.

Y por supuesto, no podemos cerrar este análisis sin mencionar el algoritmo **COMPAS** y el caso *Estado vs. Loomis* en EE. UU. Es el recordatorio definitivo de los peligros de los **algoritmos opacos**. En ese caso, la empresa propietaria se negó a revelar la metodología de cálculo, y aunque la Corte de Wisconsin validó su uso, tuvo que imponer límites severos: el algoritmo nunca puede ser la única base para una sentencia. Me pregunto, como jurista, ¿hasta qué punto es justo que una puntuación estadística, cuya lógica interna nadie comprende del todo, desempeñe siquiera un papel menor en privar a un ser humano de su libertad? La respuesta, me temo, no está en el código, sino en nuestra capacidad para mantener al humano siempre en el circuito.

Capítulo 4: Ética Algorítmica y Desafíos de la Opacidad (La "Caja Negra")

Al abordar la ética en la Inteligencia Artificial, solemos caer en el error de pensar que estamos ante un debate meramente filosófico o abstracto, cuando en realidad nos jugamos los cimientos mismos de nuestra arquitectura jurídica. No estamos simplemente ante máquinas que procesan datos, sino ante sistemas que reflejan, y a menudo amplifican, las prioridades y prejuicios de quienes los diseñan. Por ello, para un operador judicial, la ética no es un complemento opcional; es la brújula

dinámica que debe guiar la aceptación o el rechazo de estas tecnologías en los estrados.

4.1. Marco ético de la UNESCO: Una reflexión normativa necesaria

La UNESCO propone abordar la ética de la IA no como un manual estático, sino como una **reflexión normativa sistemática** que evoluciona junto a la tecnología. Este marco se asienta sobre pilares innegociables: la dignidad humana, el bienestar y la prevención de daños. Me parece fundamental recalcar que estos principios, aunque no tengan el carácter vinculante de una ley, ofrecen la base necesaria para construir regímenes regulatorios sólidos que protejan los **derechos fundamentales**. Es curioso observar cómo organizaciones como la OCDE o la Comisión Europea coinciden en que una IA confiable debe ser, ante todo, inclusiva y diversa, evitando que el progreso técnico deje atrás a los grupos más vulnerables.

4.2. El problema de la "Caja Negra" e IA Explicable (XAI)

Entramos ahora en uno de los conceptos más espinosos de este informe: la **"caja negra"**. Con este término nos referimos a sistemas cuya lógica interna es tan compleja que resulta inherentemente opaca, incluso para sus propios desarrolladores. Para un juez, trabajar con un algoritmo opaco es, permítanme la expresión, como dictar sentencia a ciegas. Si el sistema no puede explicar por qué llegó a una conclusión o cómo interrelacionó los datos, se vuelve imposible detectar salidas defectuosas o sesgadas.

Frente a esta oscuridad surge la **IA Explicable (XAI)**, que busca desarrollar modelos capaces de justificar sus decisiones. No se trata solo de transparencia técnica, sino de que la explicación sea comprensible para el usuario individual, reflejando fielmente el proceso que generó el resultado. En el ámbito judicial, la capacidad de discernir la estructura del sistema es un requisito previo para que una decisión automatizada pueda ser siquiera considerada válida.

4.3. Sesgos algorítmicos y discriminación: El espejo de nuestras faltas

Es un mito peligroso creer que los algoritmos son neutrales por ser matemáticos. El **sesgo de IA** es una diferencia sistemática en el tratamiento de ciertos grupos que, desgraciadamente, suele reproducir estereotipos y prejuicios humanos. Los sesgos pueden filtrarse en cualquier fase: desde el **sesgo de muestra** con datos no representativos (como modelos de reconocimiento facial que fallan estrepitosamente con personas de piel oscura) hasta el **sesgo de asociación**, que perpetúa desigualdades históricas al correlacionar, por ejemplo, ciertas profesiones exclusivamente con un género.

Lo que más me preocupa como redactor es el **sesgo de automatización**. Hablo de esa tendencia humana, casi instintiva, a aceptar la solución de la máquina sin someterla a crítica. En un juzgado desbordado de trabajo, la tentación de confiar ciegamente en una puntuación de riesgo es enorme, pero ceder a ella significaría abdicar de la función judicial. Debemos ser conscientes de que incluso un código

aparentemente inofensivo puede convertirse en un "monstruo racista", como ocurrió con el chatbot Tay de Microsoft, si se le alimenta con la peor cara de nuestra interacción social.

4.4. Tecnologías de alto riesgo: Biometría y Falsedades Profundas

Para cerrar este capítulo, debemos elevar la guardia ante las tecnologías de **niveles de riesgo** inaceptable o alto. La identificación biométrica y el reconocimiento facial remoto en tiempo real son herramientas extremadamente intrusivas que pueden derivar en una vigilancia universal. El riesgo de identificación incorrecta y el perfilado por raza o género atentan directamente contra el derecho a la privacidad y la libertad de movimiento.

Por otro lado, el fenómeno de los **deepfakes** o falsedades profundas plantea un desafío probatorio sin precedentes. Estas imágenes o videos generados mediante Redes Generativas Antagónicas (GAN) son capaces de crear ecosistemas enteros de desinformación que parecen auténticos. La posibilidad de que un juez acepte por error una prueba fabricada pone en jaque la presunción de inocencia y el derecho a un juicio justo. Ante este panorama, la conclusión técnica es clara: la tecnología solo debe implementarse si existen salvaguardias que mantengan siempre al **humano en el circuito**, garantizando que la última palabra no la tenga un proceso estadístico, sino una conciencia ética.

Capítulo 5: Impacto de la IA en los Derechos Humanos y Garantías Procesales

Al llegar a este punto del informe, debemos reconocer una verdad incómoda: la fuerte correlación que existe entre la salud de nuestras democracias y la integridad de nuestro sistema judicial se está viendo puesta a prueba por la tecnología. No es una exageración decir que el Estado de derecho depende de un poder judicial independiente que actúe como baluarte de los **derechos fundamentales**. Sin embargo, cuando introducimos sistemas de IA en este delicado equilibrio, corremos el riesgo de que la eficiencia opaque la justicia. Los derechos humanos no son meras sugerencias éticas; son obligaciones jurídicamente vinculantes que los Estados deben defender, incluso —o especialmente— en el entorno digital.

5.1. El debido proceso y el derecho a un juicio justo

El derecho a un juicio justo, consagrado en el artículo 14 del Pacto Internacional de Derechos Civiles y Políticos, exige que toda persona sea oída por un tribunal independiente e imparcial. Pero, ¿qué sucede cuando la "imparcialidad" se delega en un algoritmo? Me preocupa profundamente el uso de herramientas de evaluación de riesgos penales. Se nos venden como instrumentos para eliminar el sesgo humano, pero a menudo terminan siendo **algoritmos opacos** que refuerzan prejuicios históricos bajo una apariencia de objetividad matemática.

El caso *Estado vs. Loomis* en los Estados Unidos es un recordatorio escalofriante de este dilema. Permitir que un algoritmo como COMPAS, cuya metodología es un secreto comercial propiedad de una empresa privada, influya en la privación de libertad de un ciudadano es, como poco, cuestionable desde la óptica del debido proceso. La presunción de inocencia corre un peligro real si empezamos a tratar las predicciones estadísticas de reincidencia como verdades procesales. En mi opinión, un juez nunca debe ser un mero aplicador de resultados algorítmicos; su deber es ser un evaluador crítico. Al final del día, el sesgo de automatización —esa tendencia humana a confiar ciegamente en la máquina— es quizás la mayor amenaza para la individualización de las penas.

5.2. Derecho a la privacidad y protección de datos

La privacidad es, a menudo, la puerta de entrada para el ejercicio de otros derechos. Sin ella, no hay libertad de expresión ni de asociación segura. En este mundo datificado, nuestras acciones dejan una huella digital constante que la IA utiliza para construir perfiles de comportamiento increíblemente precisos. Ya no hablamos solo de que una tienda como Target pueda predecir un embarazo antes que la propia familia; hablamos de la posibilidad de que el Estado utilice el reconocimiento biométrico remoto para una vigilancia universal.

Las tecnologías de identificación biométrica, como los sistemas SARI en Italia o el uso de Clearview AI, representan **niveles de riesgo** que la sociedad debe debatir con urgencia. Cuando una cámara en un espacio público puede identificar y rastrear a personas de forma sistemática, se produce un "efecto escalofriante" que altera nuestra autonomía. Es vital que entendamos la protección de datos no solo como una derivación de la privacidad, sino como un derecho independiente que otorga a la persona el control sobre su identidad digital. No podemos permitir que la "invisibilidad de los datos" de los grupos marginados se traduzca en una discriminación automatizada y silenciosa.

5.3. Libertad de expresión y acceso a la información

La transición de la plaza pública a las plataformas digitales ha dejado la moderación de contenidos en manos de algoritmos de IA. Estos sistemas, aunque eficientes para procesar volúmenes masivos de datos, suelen ser ciegos al contexto. He reflexionado mucho sobre el caso de "la niña del napalm" en Facebook; es el ejemplo perfecto de cómo un algoritmo, programado para detectar desnudez, puede censurar una de las imágenes históricas más potentes de la humanidad.

Las herramientas de Procesamiento de Lenguaje Natural (PLN) todavía tienen dificultades con el sarcasmo, la parodia o los matices culturales. Cuando una red social decide qué voces escuchamos basándose en el "compromiso" (*engagement*), está moldeando nuestra realidad. Además, el auge de las falsedades profundas (*deepfakes*) plantea un desafío probatorio inédito en los tribunales. Si un video puede ser alterado totalmente desde cero mediante Redes Generativas Antagónicas

(GAN), la crisis de la verdad se traslada directamente al estrado, poniendo en jaque el derecho a recibir información veraz y la confianza pública en el sistema.

5.4. Derecho a un recurso efectivo

Finalmente, nada de lo anterior importa si el ciudadano no tiene un camino claro para impugnar una decisión automatizada. El derecho a un recurso efectivo es la última línea de defensa. Sin embargo, la opacidad de los sistemas de "caja negra" hace que sea casi imposible para un acusado o un demandante entender la lógica de la decisión que le afecta. ¿Cómo puedes defenderte de algo que nadie puede explicar?

Casos como el de *Deliveroo* en Italia o el escándalo de *Robodebt* en Australia nos muestran que los algoritmos pueden causar daños masivos y sistémicos. El principio de "refutabilidad" es innegociable: las personas deben tener el derecho a obtener una decisión no automatizada y a conocer el razonamiento detrás del código. Me parece alentador ver cómo regulaciones como el RGPD en la Unión Europea empiezan a consagrar el "derecho a la explicación". No obstante, la tensión entre el secreto comercial de los desarrolladores y el debido proceso sigue siendo un nudo gordiano que los jueces tendrán que desatar caso por caso. La justicia que no puede explicarse a sí misma, simplemente, deja de ser justicia.

He redactado este capítulo analizando la profunda relación entre la arquitectura técnica de la IA y su impacto en el tejido legal de nuestra sociedad. Las referencias al Kit de herramientas de la UNESCO y a la jurisprudencia internacional (como los casos *Loomis*, *Deliveroo* y *Chubbs*) sustentan cada una de estas reflexiones, asegurando que la transición hacia una justicia tecnificada no deje atrás los valores fundamentales que juramos proteger.

Capítulo 6: Modelos de Gobernanza, Regulación y el Futuro del Estado de Derecho

Al llegar al cierre de este informe, resulta evidente que el desafío no es solo técnico o ético, sino profundamente político y regulatorio. No podemos permitir que la Inteligencia Artificial opere en un vacío legal o bajo una suerte de "lejano oeste" digital. La gobernanza de la IA debe ser dinámica; no basta con leyes estáticas, necesitamos marcos que evolucionen al ritmo de las redes neuronales. Como operadores judiciales, nuestra misión es garantizar que la tecnología se pliegue a los valores del Estado de derecho y no a la inversa. Al final del día, la legitimidad del sistema de justicia reside en su capacidad para rendir cuentas ante el ciudadano, algo que una máquina, por sofisticada que sea, no puede hacer por sí sola.

6.1. Enfoques de gobernanza basados en el riesgo

Una de las tendencias más sólidas a nivel internacional es la adopción de modelos de gobernanza proporcionales al riesgo. Me parece un enfoque sensato: no todas las aplicaciones de IA son iguales. No es lo mismo un filtro de correo no deseado

que un sistema que decide si una persona debe ir a prisión preventiva. En este sentido, el proyecto de Ley de IA de la Unión Europea se ha convertido en el referente global indiscutible, estableciendo una jerarquía clara.

Este marco clasifica los sistemas en cuatro **niveles de riesgo**. En la cúspide se encuentran los de "riesgo inaceptable", que sencillamente deben prohibirse, como la calificación social por parte de los gobiernos —ese escenario distópico donde se clasifica a los ciudadanos por su comportamiento—. Luego vienen los de "alto riesgo", donde se ubica la mayor parte de las herramientas judiciales, como la evaluación de pruebas o la administración de justicia. Estos sistemas deben superar un riguroso proceso de certificación y marcado CE antes de entrar al mercado. Es un mecanismo de control *ex ante* que busca evitar que algoritmos defectuosos terminen en el escritorio de un juez. Sin embargo, me pregunto si esta clasificación será suficiente para contener la velocidad de la innovación o si terminaremos persiguiendo sombras tecnológicas.

6.2. El principio del "Humano en el Circuito" (HITL)

Frente a la autonomía creciente de las máquinas, surge el imperativo del **"Humano en el Circuito" (HITL)**. La premisa es clara: la IA nunca debe reemplazar completamente al ser humano en la toma de decisiones finales, especialmente cuando están en juego **derechos fundamentales**. Pero ojo, que la presencia de un humano no es una receta mágica. Debemos ser extremadamente vigilantes con el **sesgo de automatización**, esa tendencia casi inconsciente a validar lo que dice el algoritmo simplemente por considerarlo "objetivo" o matemático.

Existen variaciones en este control. El modelo "humano sobre el circuito" permite una supervisión durante el diseño y la operación, mientras que el "humano dentro del circuito" implica una intervención en cada decisión individual. No obstante, en un juzgado desbordado por la carga de trabajo, el riesgo de que el juez se convierta en un mero sello de goma de una predicción estadística es real. Por ello, la regulación debe prescribir que la desviación de un juez frente a una recomendación algorítmica nunca pueda ser motivo de sanción o inspección disciplinaria. El juez debe mantener su independencia y ser el evaluador crítico último de la herramienta.

6.3. Evaluaciones de impacto: HRIA y FRAIA

Para que la gobernanza sea efectiva, necesitamos herramientas de diagnóstico. Aquí es donde entran las **Evaluaciones de Impacto en los Derechos Humanos (HRIA)**. No podemos esperar a que ocurra el daño para actuar; la debida diligencia debe ser preventiva. Un ejemplo excelente es la herramienta **FRAIA** (Derechos Fundamentales y Algoritmos) desarrollada en los Países Bajos. Se trata de un diálogo interdisciplinar donde expertos legales, científicos de datos y gestores analizan cómo un algoritmo podría afectar la equidad o la privacidad antes de su despliegue.

También me gustaría destacar el marco **HUDERAF**, propuesto por el Alan Turing Institute. Este modelo combina la gestión de riesgos con la participación de las partes interesadas. No se trata de llenar un formulario y olvidarse; es un proceso continuo que valida los daños potenciales y establece protocolos de reparación. La transparencia no es solo publicar un código fuente que pocos entienden, sino explicar de forma inteligible cómo y por qué se llegó a un resultado. Un sistema de justicia que no puede explicarse a sí mismo, simplemente, deja de ser justicia.

6.4. El rol del operador judicial: Garante del Estado de derecho

Concluyo este análisis reafirmando que, en esta era tecnológica, el rol del juez es más crucial que nunca. La IA tiene un potencial inmenso para mejorar la excelencia judicial, eliminando tareas mecánicas y facilitando el acceso a la justicia. Pero no es un oráculo. Es un sistema sociotécnico que refleja los sesgos de sus creadores y de los datos con los que se entrena.

El juez del futuro debe ser un profesional con competencias digitales que le permitan interrogar a la máquina. Debe exigir que los algoritmos sean **explicables (XAI)** y rechazar aquellos que se amparen en "secretos comerciales" para ocultar su lógica interna. La individualización de las penas y la motivación de las sentencias son baluartes que no podemos delegar. Al final, la tecnología es solo una herramienta; el garante último de la libertad y del debido proceso sigue siendo, y debe seguir siendo, un ser humano dotado de conciencia ética y lealtad a la ley. El Estado de derecho se erige en nuestras mentes, y es ahí donde debemos construir las defensas contra cualquier asomo de tiranía algorítmica.

I. Listado Exhaustivo de Fuentes Utilizadas

Este informe se ha estructurado fundamentalmente a partir del **Kit de herramientas mundial sobre la IA y el Estado de derecho para el poder judicial**, publicado por la **UNESCO en 2023**. No obstante, para dotar al texto de una visión multidisciplinar, se han integrado las siguientes referencias clave citadas en dicho manual:

1. **UNESCO (2023):** *Global Toolkit on AI and the Rule of Law for the Judiciary*. París, Francia. (Fuente principal de arquitectura y casos de estudio).
2. **UNESCO (2021):** *Recomendación sobre la Ética de la Inteligencia Artificial*. (Marco normativo global para la protección de la dignidad humana).
3. **OCDE (2019):** *Principios de la OCDE sobre la Inteligencia Artificial*. (Base para la gobernanza y los sistemas de IA confiables).
4. **Comisión Europea (2021):** *Propuesta de Reglamento por el que se establecen normas armonizadas sobre Inteligencia Artificial (Ley de IA)*. (Clasificación de niveles de riesgo).
5. **Jurisprudencia Citada:**
    - *State v. Loomis* (Corte Suprema de Wisconsin, EE. UU., 2016): Sobre el uso de algoritmos de evaluación de riesgos penales.
    - *Caso Deliveroo* (Tribunal de Bolonia, Italia, 2020): Sobre la discriminación algorítmica indirecta.
    - *Escándalo de Robodebt* (Australia, 2021): Sobre los daños masivos por decisiones automatizadas deficientes.
6. **Principios FAIR (2016):** Directrices para la gestión de datos científicos (Findable, Accessible, Interoperable, Reusable).

II. Glosario de Términos

Para asegurar una interpretación unívoca del informe, se definen los siguientes conceptos fundamentales:

- **Algoritmo:** Conjunto de reglas matemáticas e instrucciones lógicas que un sistema sigue para resolver un problema o completar una tarea.
- **Aprendizaje Automático (Machine Learning - ML):** Subcampo de la IA que utiliza técnicas estadísticas para que los sistemas "aprendan" y mejoren su rendimiento a partir de datos, sin ser programados explícitamente para cada decisión.
- **Caja Negra (Black Box):** Se refiere a la opacidad de ciertos sistemas de IA (especialmente redes neuronales profundas) cuyo procesamiento interno es tan complejo que resulta incomprensible incluso para sus desarrolladores.
- **Datificación:** El proceso de convertir diversos aspectos de la vida social y del comportamiento humano en datos digitales que pueden ser procesados algorítmicamente.

- **Deepfakes (Falsedades Profundas):** Contenido multimedia (video, audio o imagen) generado mediante IA que imita de forma hiperrealista la apariencia o voz de una persona real, con el potencial de desinformar o falsificar pruebas.
- **Humano en el Circuito (Human-in-the-loop - HITL):** Principio de diseño y gobernanza que exige que un ser humano supervise, intervenga o valide las decisiones tomadas por un sistema de IA, especialmente en contextos de alto riesgo.
- **IA Explicable (XAI):** Conjunto de métodos y técnicas que permiten que los resultados de un sistema de IA sean comprensibles y rastreables para los usuarios humanos.
- **Sesgo Algorítmico:** Error sistemático en un sistema informático que genera resultados injustos o discriminatorios, generalmente debido a datos de entrenamiento no representativos o prejuicios en el diseño del modelo.

III. FAQs (Preguntas Frecuentes)

1. ¿Puede la IA sustituir la figura del juez en la toma de decisiones? No. De acuerdo con el principio del **Estado de derecho**, la función jurisdiccional requiere una conciencia ética y una lealtad a la ley que solo un ser humano posee. La IA debe entenderse exclusivamente como una herramienta de apoyo para aumentar la eficiencia administrativa o facilitar la investigación de precedentes, manteniendo siempre al humano como evaluador crítico último.

2. ¿Cuáles son los mayores riesgos de usar IA en el sistema penal? Los riesgos principales incluyen la vulneración de la presunción de inocencia, el uso de **algoritmos opacos** que impiden el derecho a la defensa y la automatización de sesgos históricos (raciales, económicos o geográficos) presentes en los datos policiales.

3. ¿Qué es el principio FAIR y por qué es vital en justicia? Es el estándar para que los datos sean Localizables, Accesibles, Interoperables y Reutilizables. En justicia, es vital porque sin una infraestructura de datos de alta calidad y bien etiquetados, cualquier sistema de IA producirá resultados erróneos, incompletos o sesgados que podrían arruinar la vida de un ciudadano.

4. ¿Cómo afecta la IA al derecho de privacidad en el ámbito judicial? El riesgo reside en la vigilancia masiva y el perfilado. Tecnologías como el reconocimiento facial remoto o la identificación biométrica pueden crear un "efecto escalofriante" en la sociedad, limitando la libertad de expresión y asociación si el ciudadano se siente constantemente monitorizado por el sistema de justicia.

5. ¿Qué debe hacer un juez si un algoritmo le da una recomendación que le genera dudas? El juez tiene la obligación de ejercer su independencia judicial. Un resultado algorítmico es solo una inferencia estadística, no una verdad legal. Si el sistema no es capaz de ofrecer una **explicación (XAI)** clara de su razonamiento, el juez debe

priorizar las garantías procesales y, si es necesario, apartarse de la recomendación de la máquina sin miedo a represalias disciplinarias.