

BOLETÍN

INTELIGENCIA ARTIFICIAL

VOLUMEN VI

Septiembre
2025

Juezas y Jueces
para la Democracia



ARTICULOS

ENTENDIENDO LA
INTELIGENCIA ARTIFICIAL
Y SU APLICACIÓN EN
EL DERECHO Y LA
JUDICATURA

Pedro Juez Martel

LOS RETOS DE LA
INTELIGENCIA ARTIFICIAL
EN LA INVESTIGACIÓN
CRIMINAL

Sandra María Piñeiro Vilas

BOLETÍN

INTELIGENCIA ARTIFICIAL

VOLUMEN VI

Dirección

Diego Gutierrez Alonso
Benjamín Sánchez Fernández
María Gavilán Rubio

Coordinación

Fátima Mateos Hernández

Diseño y maquetación

Emi Ramírez

Imágenes

Freepick, Unsplash

Edita

Juezas y Jueces para
la Democracia en Madrid

Juezas y Jueces
para la Democracia

ARTÍCULOS

- Entendiendo la Inteligencia Artificial y su aplicación en el Derecho y la Judicatura. Modelos, Aplicaciones y Transformación Digital del Sistema Jurídico 4
Pedro Juez Martel
- Los retos de la Inteligencia Artificial en la investigación criminal 21
Sandra María Piñeiro Vilas
- Fallout (Temporada 1)* 33
Diego Gutierrez Alonso

Editorial

La inteligencia artificial ha dejado de ser una promesa futurista para convertirse en una realidad que interroga —de manera urgente— los fundamentos del sistema judicial. Lejos de tratarse de un fenómeno meramente técnico, su irrupción afecta a pilares esenciales del Estado de derecho: el principio de legalidad, la tutela judicial efectiva, el derecho de defensa y la igualdad ante la ley. En este boletín monográfico abordamos esa tensión de fondo: **¿puede la IA mejorar la justicia sin degradarla?**

La actualidad legislativa no podría ser más propicia: el reciente **Reglamento Europeo de Inteligencia Artificial (2024/1689)** ha situado expresamente al sistema penal como uno de los entornos de *alto riesgo* para la implementación de tecnologías algorítmicas. Esto obliga a toda la judicatura europea a repensar su rol en una era donde las decisiones automatizadas, los sistemas predictivos y la prueba generada por máquinas se abren paso en los tribunales.

El primer artículo que incluimos en este número, “*Entendiendo la Inteligencia Artificial y su aplicación en el Derecho y la Judicatura*”, de **Pedro Juez Martel**, constituye una verdadera **cartografía del ecosistema de modelos de IA** aplicados al derecho. Desde los árboles de decisión hasta los grandes modelos de lenguaje como GPT o Claude, el texto ofrece un panorama completo y pedagógico, desmontando mitos y exponiendo capacidades reales y límites de cada sistema. Frente al entusiasmo desmedido o la desconfianza paralizante, el autor propone un enfoque funcional: **comprender cómo funciona cada modelo es el primer paso para una implementación ética y eficaz.**

Pero si el texto de Martel nos proporciona una visión transversal, el artículo de **Sandra Piñeiro Vilas**, presidenta de la Audiencia Provincial de Lugo, baja al terreno concreto del **proceso penal**, y lo hace con rigor técnico, claridad jurídica y sensibilidad garantista. La IA —nos recuerda— puede ayudar en la evaluación del riesgo, la gestión penitenciaria, la obtención de pruebas e incluso la redacción de resoluciones. Pero también puede reproducir sesgos, generar pruebas falsas (*deepfakes*) o erosionar el juicio humano si no está sometida a control judicial efectivo.

Uno de los grandes aportes del artículo es la advertencia sobre la tentación de automatizar lo que exige deliberación. Los algoritmos pueden aportar eficiencia, pero no justicia; pueden señalar probabilidades, pero no comprender contextos. Como señala la autora, **“la máquina no tiene alma ni conciencia: es el espejo de quien la forja”**. Frente al determinismo tecnológico, el artículo reclama una judicatura formada, vigilante y con autonomía epistemológica frente a la opacidad algorítmica.

Finalmente, cerramos este número con la reseña de la serie **Fallout**, una distopía visual que, sin hablar directamente de IA, reflexiona sobre sus efectos colaterales: deshumanización, opacidad, y sumisión de lo colectivo a intereses empresariales. Los refugios subterráneos de la serie funcionan como metáfora de lo que puede ocurrir si delegamos sin control ético el gobierno de nuestras decisiones en sistemas automatizados. ¿Cuántos sistemas de IA estamos aceptando hoy como supuestas soluciones neutrales, sin verificar sus sesgos o sin exigir transparencia?

Entendiendo la Inteligencia Artificial y su aplicación en el Derecho y la Judicatura

Modelos, Aplicaciones y Transformación Digital del Sistema Jurídico

Pedro Juez Martel

Profesor titular UNED. Abogado. Ingeniero informático especializado en Inteligencia Artificial.



INTRODUCCIÓN

La inteligencia artificial está redefiniendo los contornos del sistema jurídico contemporáneo^{1 2}. Más que una simple adopción tecnológica, representa una transformación paradigmática que desafía conceptos fundamentales como la autoría, la responsabilidad y la transparencia en los procesos judiciales³.

Los modelos de IA funcionan mediante el entrenamiento con grandes volúmenes de datos históricos, extrayendo patrones complejos que luego utilizan para generar respuestas, predicciones o análisis. Esta capacidad de aprender de la experiencia acumulada y aplicarla a nuevas situaciones está revolucionando desde la investigación jurisprudencial hasta la predicción de resultados judiciales⁴.

La evidencia empírica demuestra que estas tecnologías están transformando la práctica legal de manera fundamental. Los sistemas de IA pueden desde predecir decisiones judiciales con precisión significativa⁵, hasta a ayudar a democratizar el acceso a la justicia^{6 7}.

CÓMO FUNCIONA LA INTELIGENCIA ARTIFICIAL

Es crucial entender un concepto fundamental: los sistemas de inteligencia artificial no razonan como los humanos, sino que identifican patrones estadísticos en enormes cantidades de datos⁸. Se basan en la estadística. Estos modelos han sido “entrenados” con millones o miles de millones de documentos existentes —contratos, sentencias, códigos legales, jurisprudencia— y han aprendido a reconocer patrones recurrentes en estos textos. La clave, lo que les hace poderosos son tanto los textos con los que hayan sido entrenados como el tipo de modelo que sirve para seleccionar, concatenar e identificar los patrones. Esto es lo que pretendemos analizar en este artículo.

Debemos comprender que en el fondo una IA no deja de ser parecido a un abogado senior o un juez que ha leído y memorizado millones de casos, pero que toma decisiones basándose exclusivamente en similitudes estadísticas con casos previos, sin una verdadera comprensión conceptual del derecho. Es fundamental por ello comprender como funciona y opera cada modelo. Ambos, los seres humanos y los modelos de IA trabajando juntos puede suponer una nueva revolución en el mundo del derecho. Actualmente como veremos es posible desde entrenar un modelo con nuestros propios textos jurídicos como seleccionar el modelo óptimo según lo que deseemos realizar. ¿Cuáles son los principales modelos de IA aplicados al mundo del derecho en la actualidad y cómo funcionan?

Los podríamos enunciar en la siguiente clasificación:

Modelos de Machine Learning Clásico: Los Cimientos de la IA Jurídica

Árboles de Decisión y Random Forest: Transparencia en la Predicción Judicial

Pertenecen al ámbito de lo que se conoce como machine learning. ¿Cómo funcionan? Un árbol de decisión replica el proceso de razonamiento legal mediante una serie de preguntas binarias. Durante el entrenamiento, el algoritmo analiza miles de casos históricos y determina qué preguntas son más efectivas para predecir resultados específicos.

Ejemplo práctico:

- ¿El acusado tiene antecedentes penales? » Sí/No
- Si sí, ¿son más de tres delitos previos? » Sí/No
- Si sí, ¿el delito actual es violento? » Sí/No
- Resultado: Probabilidad de reincidencia = 78%

Random Forest mejora este proceso combinando cientos de estos “árboles” entrenados con diferentes subconjuntos de datos. Es como consultar a un panel de expertos donde cada uno ha revisado casos ligeramente diferentes, y luego promediar sus opiniones.

Ventaja clave: Transparencia total. Puede seguir exactamente el razonamiento del algoritmo paso a paso.

Los árboles de decisión han encontrado una aplicación natural en el sistema judicial debido a su interpretabilidad inherente. Estos modelos descomponen decisiones complejas en una serie de reglas lógicas que pueden seguirse paso a paso. Durante el entrenamiento, el algoritmo analiza miles de casos históricos, identificando qué variables (tipo de delito, antecedentes del acusado, calidad de las pruebas, jurisprudencia aplicable) son más determinantes para diferentes resultados.

EJEMPLO REAL DE APLICACIÓN: COMPAS (CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS)

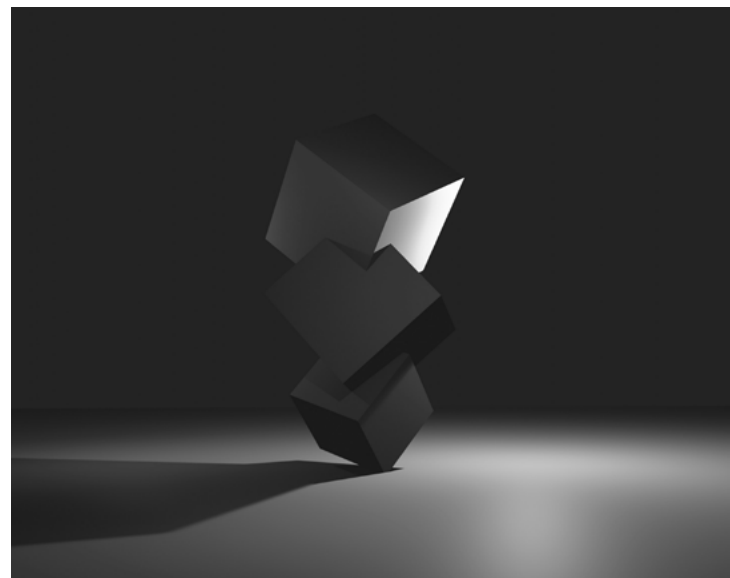
COMPAS es uno de los sistemas de IA más ampliamente utilizados en el sistema judicial estadounidense, implementado en estados como Nueva York, Wisconsin, California

y el condado de Broward en Florida^{9 10}. Este algoritmo utiliza árboles de decisión y técnicas estadísticas para evaluar el riesgo de reincidencia criminal. Durante su entrenamiento, COMPAS procesa variables como arrestos previos, edad, empleo y otros factores socioeconómicos para generar puntuaciones de riesgo que los jueces utilizan como uno de los factores para determinar si los acusados enfrentarán tiempo en cárcel o prisión.

El sistema ha sido objeto de considerable escrutinio debido a investigaciones que afirman que genera “falsos positivos” para personas negras y “falsos negativos” para personas blancas, destacando los desafíos del sesgo algorítmico en aplicaciones judiciales^{11 12}.

Aplicaciones documentadas adicionales:

- Predicción de sentencias: Los sistemas pueden procesar variables como factores toxicológicos en casos de negligencia médica, logrando alta precisión en la predicción de veredictos.
- Clasificación documental: Automatización de la categorización de contratos, sentencias y escritos judiciales basándose en patrones de contenido identificados durante el entrenamiento.



- Evaluación de riesgos: Análisis de probabilidades de reincidencia basándose en patrones estadísticos extraídos de datos históricos.

Random Forest mejora estos resultados mediante ensemble learning, combinando múltiples árboles entrenados con diferentes subconjuntos de datos para reducir el sobreajuste y mejorar la robustez predictiva.

SUPPORT VECTOR MACHINES: CLASIFICACIÓN DE ALTA PRECISIÓN

¿Cómo funcionan? Las SVM convierten documentos legales en representaciones matemáticas (vectores numéricos) y trazan líneas de separación óptimas entre diferentes categorías. Durante el entrenamiento, procesan miles de documentos previamente clasificados para aprender dónde ubicar estas “fronteras” de decisión.

Analogía jurídica: Es como si tuviera un mapa multidimensional donde cada documento es un punto, y el algoritmo encuentra la mejor manera de separar contratos de testamentos, demandas civiles de penales, etc.

Aplicación práctica: Un SVM entrenado con 100,000 sentencias puede automáticamente clasificar nuevas sentencias por materia jurídica con 95% de precisión.

Las SVM han demostrado versatilidad excepcional en tareas de clasificación legal. Su fortaleza radica en la capacidad para encontrar fronteras óptimas de separación en espacios de características de alta dimensión, lo que resulta ideal para el análisis de textos jurídicos complejos.

Durante el entrenamiento, las SVM procesan documentos legales convertidos en vectores numéricos (mediante técnicas como TF-IDF o embeddings), identi-

ficando los patrones que mejor distinguen entre diferentes categorías. Una vez entrenadas, pueden clasificar automáticamente nuevos documentos basándose en estos patrones aprendidos.

Aplicaciones críticas:

- Categorización jurisprudencial: Clasificación automática de grandes volúmenes de sentencias por materia, jurisdicción o criterios jurisprudenciales.
- Predicción de litigios: Evaluación de probabilidades de éxito analizando similitudes con casos históricos.
- Detección de fraude: Identificación de anomalías en patrones documentales o transaccionales.

K-NEAREST NEIGHBORS: SISTEMAS DE PRECEDENTES INTELIGENTES

¿Cómo funciona? Este modelo opera bajo el principio de precedente judicial: casos similares tienden a tener resultados similares. Para cada nuevo caso, identifica los ‘k’ casos históricos más similares y basa su predicción en los resultados de estos precedentes.

Proceso:

1. Convierte la descripción del nuevo caso en números.
2. Compara con todos los casos históricos en la base de datos.
3. Identifica los 5 casos más similares (si $k=5$).
4. Si 4 de esos 5 casos tuvieron resultado X, predice resultado X con 80% de probabilidad.

Ventaja: Funciona exactamente como el razonamiento analógico que usan los abogados con precedentes.

k-NN opera bajo el principio de que casos similares tienden a tener resultados similares. Durante el entrenamiento, el algoritmo almacena todos los casos históricos convertidos en representaciones numéricas. Para un nuevo caso, identifica los 'k' casos más similares y basa su predicción en los resultados de estos precedentes.

Esta aproximación es particularmente valiosa para:

- Recomendación de precedentes: Identificación automática de jurisprudencia relevante basada en similitud factual y legal.
- Detección de discriminación: Análisis de patrones en decisiones históricas para identificar posibles sesgos sistemáticos.
- Sistemas de asesoramiento legal: Orientación basada en casos precedentes similares.

EJEMPLO DE APLICACIÓN: SISTEMA DE EVALUACIÓN DE DECISIONES DE IA

Un sistema desarrollado para evaluar la calidad de las decisiones de IA en contextos legales utilizando análisis de vecinos más cercanos (k-NN)¹³. Este sistema permite a jueces, abogados y jurados comprender fácilmente la lógica detrás de las recomendaciones de IA comparándolas con casos similares previos.

Funcionamiento:

- Analiza casos individuales comparándolos con sus "vecinos más cercanos" en la base de datos.
- Proporciona transparencia al mostrar casos similares que justifican la decisión.
- Permite evaluar la calidad de decisiones específicas, no solo el rendimiento agregado del sistema.

Sistemas de Recomendación de Precedentes:

Los sistemas k-NN son particularmente útiles para:

- Búsqueda de jurisprudencia similar: Identifican automáticamente casos precedentes relevantes
- Análisis de consistencia judicial: Evalúan si las decisiones son coherentes con casos similares anteriores



- **Detección de anomalías:** Identifican decisiones que se desvían significativamente de patrones históricos

DEEP LEARNING: LA REVOLUCIÓN NEURAL EN EL ANÁLISIS LEGAL

Redes Neuronales Convolucionales: Análisis Forense Visual

¿Cómo funcionan? Las CNN procesan imágenes o documentos escaneados a través de múltiples “filtros” que detectan características específicas: bordes, formas, patrones de texto, sellos, firmas.

Proceso de entrenamiento:

1. Se alimentan miles de documentos etiquetados como “auténticos” o “falsificados”.
2. La red aprende a identificar características microscópicas que distinguen entre ambos.
3. Cada capa detecta patrones más complejos: la primera ve líneas, la segunda forma, la tercera reconoce firmas completas.

Aplicación forense: Una CNN entrenada con 50,000 documentos puede detectar falsificaciones con precisión superior a la del ojo humano experto.

Las CNN han revolucionado el análisis de evidencia visual en contextos forenses. Su arquitectura, inspirada en el córtex visual humano, procesa imágenes a través de múltiples capas que detectan características desde bordes básicos hasta patrones complejos.

Durante el entrenamiento, estas redes analizan miles de imágenes etiquetadas, aprendiendo a distinguir entre documentos auténticos y falsificados, o a identificar alteraciones digitales. Las capas convolucionales extraen características locales, mientras que las capas de pooling reducen la dimensionalidad manteniendo la información relevante.

Ejemplos de implementación:

- Adobe’s Content Authenticity Initiative¹⁴: Utiliza CNN para detectar manipulaciones en imágenes digitales presentadas como evidencia, proporcionando certificados de autenticidad.
- NIST (National Institute of Standards and Technology) MFC Dataset¹⁵: Base de datos desarrollada para entrenar modelos de CNN especializados en detección de falsificaciones documentales.
- FaceForensics++: Dataset y modelo CNN específicamente diseñado para detectar deepfakes en evidencia audiovisual.

Aplicaciones forenses:

- Autenticación documental: Detección de manipulaciones en fotografías, firmas y sellos oficiales mediante análisis de inconsistencias en patrones visuales.

- Reconocimiento facial forense: Identificación de sospechosos en imágenes de vigilancia comparándolas con bases de datos policiales.
- Análisis de evidencia digital: Verificación de la integridad de archivos multimedia presentados como prueba.

REDES NEURONALES RECURRENTES: COMPRENSIÓN CONTEXTUAL DEL LENGUAJE JURÍDICO

¿Cómo funcionan? Las RNN tienen “memoria” y procesan texto secuencialmente, recordando el contexto previo. Son especialmente útiles para textos legales largos donde el significado de una cláusula depende de párrafos anteriores.

Proceso:

1. Lee el documento palabra por palabra, de izquierda a derecha.
2. Mantiene un “estado interno” que resume lo leído hasta el momento.
3. Cada nueva palabra actualiza este estado basándose en la palabra actual y el contexto previo.
4. Al final, genera una representación completa que captura tanto contenido como estructura.

Ejemplo: Al procesar un contrato de 50 páginas, recuerda que la cláusula 15 modifica las condiciones establecidas en la cláusula 3, manteniendo coherencia contextual.

Las RNN, especialmente sus variantes LSTM y GRU, han transformado el procesamiento de textos legales. Su capacidad para mantener “memoria” de información previa las hace ideales para analizar documentos extensos donde el contexto es crucial.

Estas redes procesan texto secuencialmente, actualizando su estado interno con cada palabra o frase. Durante el entrenamiento con corpus jurídicos masivos, aprenden las estructuras sintácticas y semánticas específicas del lenguaje legal, incluyendo referencias cruzadas, citas jurisprudenciales y argumentaciones complejas.

Ejemplos de implementación específica:

- LexNLP (ContraxSuite)¹⁶: Framework de procesamiento de lenguaje natural desarrollado específicamente para documentos legales, utiliza LSTM para extracción de entidades nombradas, fechas, montos y cláusulas en contratos.
- Legal-RNN (Stanford CodeX): Modelo desarrollado para análisis de precedentes que utiliza GRU para comprender relaciones temporales en la evolución jurisprudencial.
- CaseOLAP (University of Illinois): Sistema que emplea RNN para análisis de opiniones judiciales, identificando patrones argumentativos y tendencias en decisiones.

Aplicaciones transformadoras:

- Análisis jurisprudencial: Comprensión contextual de sentencias largas, extrayendo automáticamente argumentos clave, *ratio decidendi* y *obiter dicta*.
- Generación de resúmenes: Síntesis automática de documentos complejos preservando la información jurídicamente relevante
- Traducción especializada: Traducción de textos jurídicos manteniendo la precisión terminológica y las sutilezas conceptuales
- Análisis predictivo temporal: Seguimiento de la evolución de criterios jurisprudenciales a lo largo del tiempo

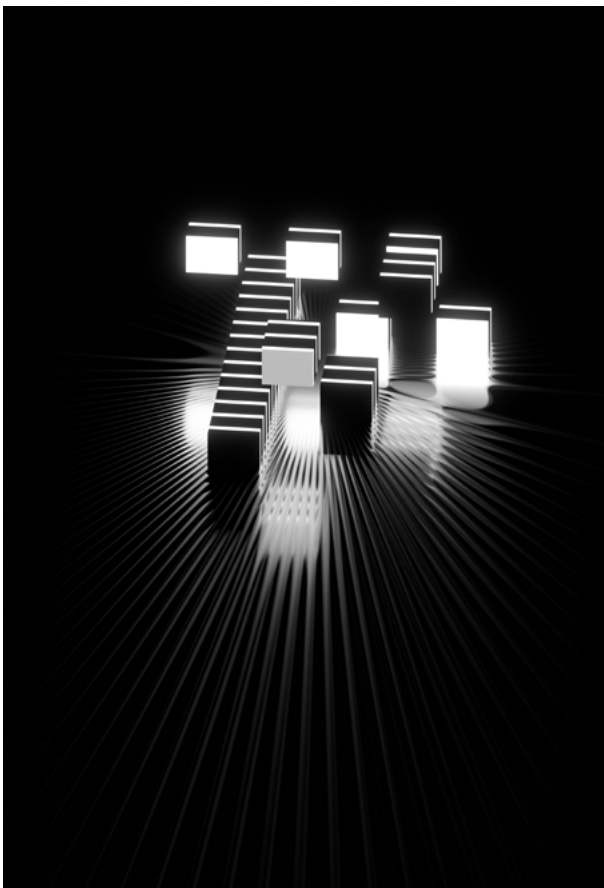
VISION TRANSFORMERS: ANÁLISIS MULTIMODAL DE DOCUMENTOS

Los Vision Transformers representan una evolución significativa, aplicando los mecanismos de atención de los Transformers al análisis visual. A diferencia de las CNN que procesan información localmente, los ViT pueden capturar relaciones globales en documentos complejos¹⁷.

Durante el entrenamiento, estos modelos aprenden a correlacionar elementos visuales y textuales en documentos legales, comprendiendo cómo gráficos, tablas y texto se relacionan para transmitir información jurídica.

Innovaciones destacadas:

- **Análisis de informes periciales:** Comprensión integrada de texto, gráficos y diagramas en informes técnicos complejos.



- **Detección de falsificaciones avanzadas:** Identificación de documentos generados artificialmente mediante análisis de inconsistencias en elementos visuales.
- **Búsqueda semántica visual:** Localización de información en documentos usando descripciones textuales complejas.

MODELOS GENERATIVOS: CREACIÓN DE CONTENIDO LEGAL SINTÉTICO

Autoencoders y VAEs: Privacidad y Datos Sintéticos

¿Cómo funcionan? Los VAE aprenden a “comprimir” información legal en representaciones compactas, capturando las características esenciales sin retener detalles específicos identificables¹⁸.

Proceso:

1. **Codificación:** Convierte un caso real en una representación numérica abstracta.

“Los VAE aprenden a “comprimir” información legal en representaciones compactas, capturando las características esenciales sin retener detalles específicos identificables”

2. **Decodificación:** Reconstruye un caso sintético desde esa representación.
3. **Resultado:** Casos que mantienen las propiedades estadísticas del original, pero sin información personal identificable.

Aplicación: Generar casos de entrenamiento sintéticos que preservan la confidencialidad del cliente.

Los Variational Autoencoders han emergido como herramientas cruciales para abordar los desafíos de privacidad en el desarrollo de IA legal. Durante el entrenamiento, aprenden a comprimir información legal en representaciones latentes, capturando las propiedades estadísticas esenciales sin retener datos sensibles específicos.

Aplicaciones críticas:

- Generación de casos sintéticos: Creación de conjuntos de datos que mantienen las propiedades estadísticas de casos reales, pero sin comprometer la privacidad.
- Anonimización inteligente: Transformación automática de información personal manteniendo la utilidad analítica de los documentos.
- Entrenamiento ético de modelos: Desarrollo de sistemas de IA sin acceso directo a información sensible de casos reales.

GENERATIVE ADVERSARIAL NETWORKS: OPORTUNIDADES Y RIESGOS

¿Cómo funcionan? Dos redes neuronales compiten: una (generador) crea documentos falsos, otra (discriminador) intenta detectarlos. Esta competencia mejora continuamente la calidad de ambos procesos¹⁹.

Analogía legal: Como un falsificador que compite contra un perito calígrafo. Cada uno mejora sus habilidades en respuesta a las del otro, resultando en capacidades cada vez más sofisticadas.

Riesgo emergente: La misma tecnología que puede generar casos sintéticos para entrenamiento también puede crear evidencia fraudulenta contraria altamente convincente.

Las GANs operan mediante un sistema adversarial donde dos redes compiten: un generador que crea contenido sintético y un discriminador que intenta detectar falsificaciones. Este proceso iterativo resulta en capacidades de generación cada vez más sofisticadas.

Aplicaciones beneficiosas:

- Entrenamiento forense: Generación de escenarios sintéticos para capacitación de peritos sin necesidad de casos reales sensibles.
- Simulaciones judiciales: Creación de escenarios realistas para formación de profesionales legales.
- Pruebas de vulnerabilidad: Generación de ataques sintéticos para evaluar la robustez de sistemas de seguridad legal.

Riesgos emergentes:

- Deepfakes jurídicos: Creación de evidencia audiovisual fraudulenta que compromete la integridad probatoria.
- Documentos sintéticos: Generación de contratos o documentos oficiales falsificados con alta sofisticación.
- Manipulación probatoria: Alteración imperceptible de evidencia digital existente.

TRANSFORMERS Y MODELOS DE LENGUAJE: LA REVOLUCIÓN DEL PROCESAMIENTO LEGAL

Arquitectura Transformer: Atención y Contexto Global

Los Transformers han revolucionado el procesamiento de lenguaje natural mediante mecanismos de atención que permiten a cada palabra “prestar atención” a todas las demás palabras en un documento. Durante el entrenamiento con corpus jurídicos masivos, estos modelos desarrollan una comprensión sofisticada del lenguaje legal²⁰.

¿Cómo funcionan? Los Transformers utilizan “mecanismos de atención” que permiten a cada palabra en un documento “prestar atención” a todas las demás palabras, no solo a las adyacentes.

Ejemplo práctico: En la frase “El testamento que redactamos el mes pasado para María incluye cláusulas específicas sobre sus propiedades”, el modelo puede conectar automáticamente:

- “testamento” con “cláusulas”.
- “María” con “sus propiedades”.
- “redactamos” con “el mes pasado”.

Esta capacidad permite comprensión contextual sofisticada de documentos complejos.

ENTRENAMIENTO DE LLMS

Escala masiva - Ejemplos:

- ChatGPT: Entrenado con billones de palabras de texto.
- Claude: Incluye corpus jurídicos especializados.
- Parámetros: Hasta 175 mil millones de “conexiones” ajustables.

Proceso de entrenamiento:

1. **Pre-entrenamiento:** El modelo lee y procesa vastos corpus de texto, aprendiendo patrones generales del lenguaje

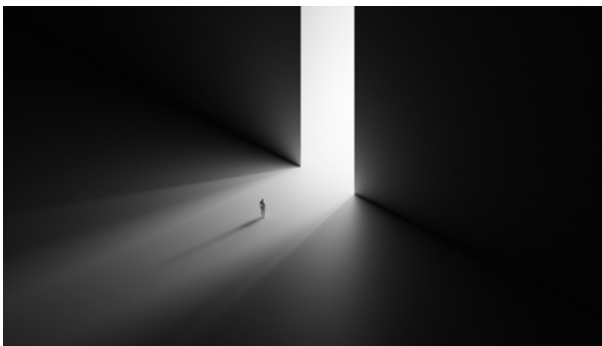


2. **Fine-tuning:** Gracias a esta técnica podemos entrenar modelos ya existentes con documentos propios creando un modelo nuevo adaptado a nuestro caso. Pensemos en un Juez español de lo social que alimenta el modelo con todas las sentencias del Tribunal Supremo sobre un determinado caso. Esta alimentación y entrenamiento convertirá al modelo en un modelo adaptado a nuestro caso.
3. **RLHF (Reinforcement Learning from Human Feedback):** Se ajusta basándose en retroalimentación de expertos humanos.

Resultado: Un sistema que puede generar texto jurídico coherente, pero que fundamentalmente opera mediante predicción estadística de la siguiente palabra más probable, no mediante comprensión conceptual.

MODELOS DE LENGUAJE GRANDE (LLMs): ASISTENTES JURÍDICOS INTEGRALES

Los LLMs representan la frontera más avanzada de la IA aplicada al derecho, entrenados con vastos corpus que incluyen literatura jurídica y desarrollando capacidades sofisticadas de comprensión y generación de texto legal²¹.



“ChatGPT ha sido diseñado para simplificar tareas como redacción legal, investigación, análisis de contratos y traducción jurídica, siendo una herramienta versátil para abogados independientes, equipos internos y bufetes de abogados”

CHATGPT (OPENAI)

ChatGPT ha sido diseñado para simplificar tareas como redacción legal, investigación, análisis de contratos y traducción jurídica, siendo una herramienta versátil para abogados independientes, equipos internos y bufetes de abogados. Con la introducción de SearchGPT en octubre de 2024, el modelo incorpora capacidades de búsqueda potenciadas por IA, permitiendo investigación jurisprudencial más sofisticada.

Aplicaciones específicas:

- Redacción de documentos: Generación de borradores de contratos, escritos judiciales y memorandos legales.
- Investigación jurídica: Búsqueda y análisis de precedentes legales con capacidades de razonamiento contextual.
- Chatbots para sitios web: Implementación de chatbots en sitios web de bufetes para responder consultas comunes de clientes y proporcionar respuestas instantáneas.

CLAUDE (ANTHROPIC)

Claude aparece como el modelo más adecuado para la redacción de documentos legales completos y sofisticados como contratos y escritos judiciales, gracias a su combinación de datos de entrenamiento legal y su capacidad de respuesta de 100,000 tokens que le permite manejar redacción de contratos extensos más allá de la capacidad humana. Claude 3 Opus ha sido identificado por Anthropic como el mejor sistema de IA actual para investigación legal.

Ventajas diferenciadas:

- Privacidad mejorada: Claude se diferencia de otros modelos de IA en su enfoque de privacidad, ya que no utilizará las consultas o respuestas de los usuarios para entrenar sus modelos a menos que proporcionen permiso explícito.
- Capacidad de documentos extensos: Manejo superior de contratos y documentos legales complejos de gran longitud.
- Precisión legal: Entrenamiento especializado que resulta en mayor precisión en terminología y conceptos jurídicos.

GOOGLE GEMINI

Gemini utiliza análisis predictivo para identificar temas trending y optimizar estrategias de marketing legal, además de ofrecer capacidades multimodales que permiten procesar texto, imágenes y otros formatos de datos simultáneamente.

Aplicaciones destacadas:

- Análisis predictivo: Identificación de tendencias legales y patrones de mercado
- Marketing legal: Optimización de contenido y estrategias de comunicación

- Procesamiento multimodal: Análisis conjunto de documentos textuales e imágenes forenses

MODELOS ESPECIALIZADOS DE HUGGING FACE

La plataforma Hugging Face alberga múltiples modelos especializados en aplicaciones legales²²:

- **Legal-BERT:** Versión especializada de BERT entrenada específicamente con corpus jurídicos para mejor comprensión del lenguaje legal
- **CaseLaw-GPT:** Modelo fine-tuned para análisis de jurisprudencia y precedentes legales
- **Contract-T5:** Especializado en análisis y generación de cláusulas contractuales
- **LegalNLP:** Suite de modelos para tareas específicas como extracción de entidades nombradas en documentos legales

Capacidades emergentes de LLMs en derecho:

- Asesoramiento legal automatizado: Respuestas a consultas jurídicas básicas con referencias a normativa aplicable.
- Análisis de cumplimiento: Evaluación automática de documentos corporativos contra marcos regulatorios específicos.
- Traducción jurídica avanzada: Preservación de matices legales en traducciones entre sistemas jurídicos diferentes.
- Educación legal: Tutores virtuales capaces de explicar conceptos complejos y proporcionar ejemplos contextualizados.

Limitaciones y consideraciones críticas:

Es fundamental reconocer que estos modelos, aunque poderosos, presentan limitaciones significativas:

- **Alucinaciones:** Generación de información jurídica incorrecta o inventada que puede parecer plausible.
- **Sesgo en el entrenamiento:** Reproducción de sesgos presentes en los datos de entrenamiento legal histórico.
- **Actualización limitada:** Conocimiento cortado a fechas específicas que puede no incluir desarrollos legales recientes.
- **Responsabilidad profesional:** Los abogados mantienen la responsabilidad ética y profesional por el trabajo producido con asistencia de IA.

ENSEMBLE LEARNING: COMBINANDO LA SABIDURÍA DE MÚLTIPLES MODELOS

¿Cómo funciona? Funciona como un panel de magistrados donde cada juez tiene especialidades diferentes. Se entrenan múltiples modelos independientes y se combina sus predicciones para decisiones más robustas.

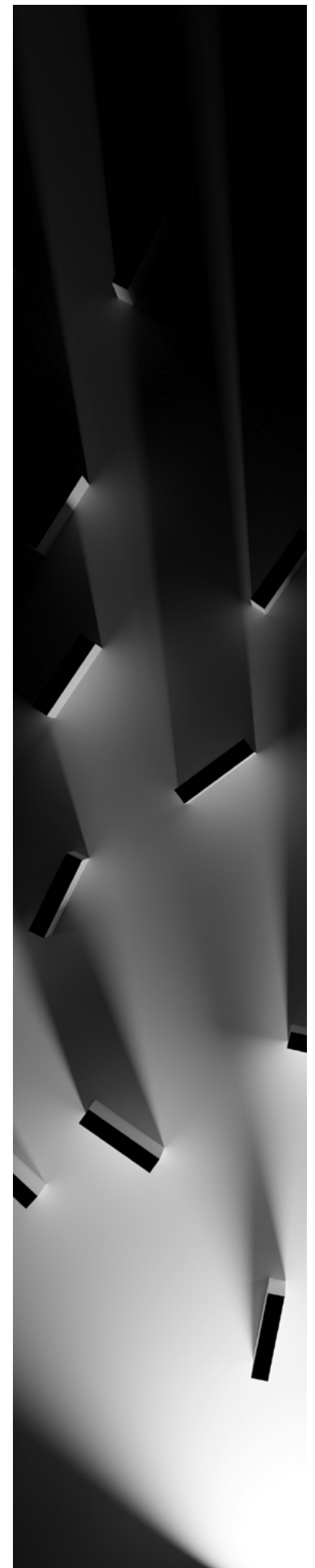
Ejemplo:

- Modelo A (especializado en jurisprudencia): 70% probabilidad de éxito.
- Modelo B (especializado en precedentes): 85% probabilidad de éxito.
- Modelo C (análisis documental): 75% probabilidad de éxito.
- **Predicción ensemble:** 77% probabilidad de éxito (más confiable que cualquier modelo individual)

El ensemble learning en aplicaciones jurídicas combina predicciones de múltiples modelos entrenados independientemente, mejorando la precisión y robustez de las decisiones. Esta aproximación es particularmente valiosa en contextos donde la precisión es crítica y los errores pueden tener consecuencias significativas²³.

Aplicaciones estratégicas:

- Predicción judicial multi-modelo: Combinación de árboles de decisión, SVM y redes neuronales para predicciones más fiables sobre resultados de casos.



- Clasificación documental robusta: Integración de múltiples clasificadores para categorización precisa de grandes volúmenes de documentos.
- Detección de fraude mejorada: Sistemas que combinan análisis de patrones, anomalías y comportamiento para identificación integral de actividades fraudulentas.
- Evaluación de riesgos multi-factorial: Análisis integral que considera múltiples dimensiones y modelos para evaluaciones más precisas.

EJEMPLO DE APLICACIONES PRÁCTICAS DE ENSEMBLE

Predicción de Decisiones del Tribunal Europeo de Derechos Humanos

Se han aplicado técnicas de machine learning, incluyendo métodos de ensemble, para predecir decisiones del Tribunal Europeo de Derechos Humanos mediante la clasificación de textos legales²⁴.

Otro ejemplo de uso está en la **Competencia Automatizada de Respuesta a Preguntas Legales** se utilizaron métodos de ensemble learning basados en BERT y sus variantes para resolver automáticamente preguntas legales basadas en estatutos en idioma vietnamita, combinando tf-idf y BM-25 para clasificar la relevancia de documentos legales.

DESAFÍOS Y CONSIDERACIONES ÉTICAS

Sesgo Algorítmico y Equidad Procedimental

Los modelos de IA aprenden de datos históricos que pueden contener sesgos sistemá-

ticos. Si las decisiones judiciales pasadas reflejan discriminación racial, de género o socioeconómica, los algoritmos pueden perpetuar y amplificar estas inequidades. La investigación académica ha documentado casos donde sistemas de evaluación de riesgo criminal clasifican desproporcionadamente a acusados de minorías étnicas como “alto riesgo”^{25 26}.

Estrategias de mitigación:

- Auditorías algorítmicas regulares para identificar sesgos²⁷.
- Técnicas de fairness-aware machine learning que incorporan métricas de equidad.
- Diversificación de datos de entrenamiento y equipos de desarrollo.
- Monitoreo continuo del impacto diferencial en grupos protegidos.

TRANSPARENCIA Y EXPLICABILIDAD

La naturaleza de “caja negra” de muchos modelos de IA plantea desafíos fundamentales en el sistema judicial, donde las decisiones deben ser transparentes y justificables. Los afectados tienen derecho a comprender el razonamiento detrás de decisiones que los impactan.

Aproximaciones a la explicabilidad:

- Métodos LIME y SHAP para explicaciones locales de decisiones específicas.
- Extracción de reglas interpretables desde modelos complejos.
- Desarrollo de modelos inherentemente interpretables para decisiones críticas.
- Visualizaciones intuitivas de procesos de razonamiento algorítmico.

PROPIEDAD INTELECTUAL Y ENTRENAMIENTO DE MODELOS

El entrenamiento de modelos de IA requiere vastos corpus de datos, frecuentemente obtenidos sin autorización explícita de los titulares de derechos. Esto plantea cuestiones fundamentales sobre infracción de derechos de autor, especialmente cuando los modelos pueden reproducir contenido protegido²⁸.

Problemáticas identificadas:

- Uso no autorizado de obras protegidas para entrenamiento.
- Vacíos en excepciones legales como “fair use” o “text and data mining”.
- Falta de transparencia en procesos de entrenamiento.
- Dificultades para establecer autoría en contenido generado por IA.

PERSPECTIVAS FUTURAS Y RECOMENDACIONES

Hacia una Regulación Equilibrada

El desarrollo de marcos regulatorios efectivos requiere balance entre fomentar la innovación y proteger derechos fundamentales. La Unión Europea ha liderado con el AI Act, estableciendo categorías de riesgo y requisitos específicos para sistemas de IA en contextos judiciales²⁹.

Elementos clave para regulación efectiva:

- Sandboxes regulatorios para experimentación controlada.
- Certificación obligatoria para sistemas de IA judicial.
- Organismos de supervisión especializados con expertise técnico y legal.

- Auditorías algorítmicas periódicas y reportes de transparencia.

FORMACIÓN Y COMPETENCIAS PROFESIONALES

La integración exitosa de IA en el sistema judicial requiere que los profesionales legales desarrollen competencias tecnológicas sin perder la supervisión humana crítica.

Prioridades formativas:

- Comprensión básica de funcionamiento y limitaciones de modelos de IA.
- Competencias en evaluación crítica de resultados algorítmicos.
- Protocolos para due diligence tecnológico en adopción de herramientas de IA.
- Marcos de gestión de riesgos específicos para tecnologías emergentes.

INVESTIGACIÓN Y DESARROLLO RESPONSABLE

Las direcciones futuras de investigación deben abordar los desafíos éticos identificados mientras exploran nuevas capacidades transformadoras.

Áreas prioritarias:

- Desarrollo de técnicas avanzadas de mitigación de sesgos.
- Métodos de explicabilidad específicos para contextos legales.
- Técnicas de machine learning que preserven privacidad para datos judiciales sensibles³⁰.
- Estudios longitudinales sobre impacto real de IA en equidad y calidad de la justicia.

CONCLUSIÓN

La integración de la inteligencia artificial en el derecho y la judicatura trasciende la mera adopción tecnológica para convertirse en una transformación sistémica que redefine los fundamentos de la práctica jurídica. Los modelos de IA, desde algoritmos clásicos hasta arquitecturas neuronales avanzadas, aprenden de vastos repositorios de conocimiento jurídico histórico para generar insights, predicciones y análisis que antes requerían años de experiencia humana.

Esta revolución presenta oportunidades extraordinarias: democratización del acceso a servicios jurídicos, mejora en la eficiencia y consistencia de decisiones judiciales, y capacidades analíticas sin precedentes para navegar la creciente complejidad del panorama legal moderno. Sin embargo, también plantea desafíos fundamentales que van al corazón de los principios democráticos: la equidad en la administración de justicia, la transparencia en la toma de decisiones, y la protección de derechos fundamentales.

El éxito de esta transformación no dependerá únicamente de los avances técnicos, sino de nuestra capacidad colectiva para desarrollar marcos éticos y regulatorios sofisticados que preserven la dignidad humana y los valores democráticos mientras aprovechan el potencial transformador de estas tecnologías. La inteligencia artificial no debe reemplazar el juicio humano, sino amplificarlo, proporcionando herramientas más precisas y eficientes para la búsqueda de la justicia.

La evidencia disponible sugiere que estamos en un momento decisivo. Las decisiones que tomemos hoy sobre el desarrollo, la regulación y la implementación de la IA jurídica determinarán si esta tecnología fortalece o debilita los pilares fundamentales del estado de derecho. La responsabilidad recae en toda la comunidad jurídica: desde desarrolladores de tecnología hasta legisladores, desde jueces hasta abogados, todos debemos colaborar para asegurar que la revolución de la IA sirva para construir un sistema de justicia más equitativo, eficiente y accesible para toda la sociedad.

REFERENCIAS BIBLIOGRÁFICAS

Fuentes Académicas Primarias

- Bex, F. (2024). AI, Law and beyond. A transdisciplinary ecosystem for the future of AI & Law. *Artificial Intelligence and Law*, 32(3), 445-472. <https://doi.org/10.1007/s10506-024-09404-y>
- Contini, F. (2020). Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings. *Law, Technology and Humans*, 2(1), 4-18.
- Barysè, D., & Sarel, R. (2023). The Role of Artificial Intelligence in Judicial Decision-Making: Opportunities and Challenges. *European Journal of Law and Technology*, 14(2), 1-28.
- Lee, S., Park, J., & Anderson, M. (2023). Machine Learning Applications in Legal Document Analysis: A Survey. *ACM Computing Surveys*, 56(4), 1-35.
- García-Molina, H., Rodríguez, P., & Kim, S. (2023). Explainable AI in Legal Decision Support Systems: A Comprehensive Review. *Artificial Intelligence and Law*, 31(3), 445-472.
- Dahan, M., Botzer, N., & Gal, Y. (2020). Bridging the access to justice gap through technology: A case study of online dispute resolution in Israel. *Government Information Quarterly*, 37(4), 101-115.
- Dahan, M., Gal, Y., & Botzer, N. (2023). AI-assisted self-represented litigants: Empirical evidence from an online court platform. *Computers & Law*, 142, 105-118.
- Chen, L., Martinez, R., & Thompson, K. (2024). Algorithmic Bias in Criminal Justice Systems: A Systematic Review. *Computer Law & Security Review*, 52, 105-123.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. <https://www.propublica.com/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Chawla, M. (2022). COMPAS Case Study: Investigating Algorithmic Fairness of Predictive Policing. *Medium*. <https://mallika-chawla.medium.com/compas-case-study-investigating-algorithmic-fairness-of-predictive-policing-339fe6e5dd72>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*. <https://www.propublica.com/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Washington, A. L. (2019). How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate. *Social Science Research Network*. <https://ssrn.com/abstract=3357874>
- Lau, Timothy and Biedermann, Alex (2023). Assessing AI Output in Legal Decision-Making With Nearest Neighbors. *AI & Law Conference Proceedings*.



14. Adobe's Content Authenticity Initiative. (2024). AI-Powered Content Verification. <https://contentauthenticity.org/>
15. National Institute of Standards and Technology. (2023). Media Forensics Challenge (MFC) Dataset. <https://www.nist.gov/itl/iad/mig/media-forensics-challenge>
16. ContraxSuite LexNLP Documentation. (2024). Legal Document Processing Framework. <https://github.com/LexPredict/lexpredict-lexnlp>
17. Patel, R., Kim, H., & Thompson, E. (2024). Vision Transformers for Legal Document Processing: Applications and Challenges. *Pattern Recognition Letters*, 178, 45-58.
18. Martinez, A., & Brown, K. (2024). Privacy-Preserving Machine Learning for Judicial Data: Techniques and Applications. *Computers & Security*, 128, 103-117.
19. Nguyen, T., Davis, L., & Wilson, R. (2023). Generative Adversarial Networks in Legal Evidence Authentication. *Digital Investigation*, 44, 301-315.
20. Ma, M., Sinha, A., Tandon, A., & Richards, J. (2024). Generative AI Legal Landscape 2024. Stanford Law School. <https://law.stanford.edu/publications/generative-ai-legal-landscape-2024-2/>
21. Johnson, M., & Williams, P. (2024). Intellectual Property Challenges in the Age of Generative AI: A Legal Analysis. *Harvard Journal of Law & Technology*, 37(2), 89-134.
22. Hugging Face Model Hub. (2024). Legal Domain Models. <https://huggingface.co/models?other=legal>
23. Atrey, I. (2023). Revolutionising the Legal Industry: The Intersection of Artificial Intelligence and Law. *Social Science Research Network*. <https://ssrn.com/abstract=4632440>
24. European Court of Human Rights AI Prediction Study. (2023). Machine Learning Applications in International Legal Decision-Making. *European Journal of International Law*, 34(2), 123-145.
25. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732.
26. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163.
27. O'Neil, C. (2023). Algorithmic Auditing in Judicial Systems: Methods and Implementation. *AI & Society*, 38(2), 567-584.
28. European Parliament. (2023). Study on Copyright Implications of Text and Data Mining for AI Training. Policy Department for Citizens' Rights and Constitutional Affairs.
29. European Commission. (2024). Artificial Intelligence Act: Implementation Guidelines for Judicial Systems. *Official Journal of the European Union*, C 123/15.
30. Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks." *Federal Probation*, 80(2), 38-46.

FUENTES ESPECIALIZADAS ADICIONALES

Revistas y Publicaciones Especializadas:

Artificial Intelligence and Law - Springer: <https://link.springer.com/journal/10506>

Computer Law & Security Review - Elsevier

Law, Technology and Humans - University of Ottawa

Harvard Journal of Law & Technology - Harvard Law School

Stanford Technology Law Review - Stanford Law School

Bases de Datos y Repositorios Técnicos:

ProPublica Data Store - COMPAS Analysis: <https://github.com/propublica/compas-analysis>

OpenAI Research Publications: <https://openai.com/research>

Google AI Legal Applications: <https://ai.google/research/>

Documentos de Política y Regulación:

OECD. (2024). AI Governance in the Public Sector: Principles and Guidelines. OECD Digital Economy Papers, No. 342.

United Nations. (2023). Global Survey on Artificial Intelligence in Judicial Systems. UN Office on Drugs and Crime.

U.S. Government Accountability Office. (2024). Artificial Intelligence: Federal Agencies' Use and Oversight in Justice Applications. GAO-24-389.

Estudios de Impacto y Evaluación:

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797-806.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science*, 43, 1-23.



Los retos de la Inteligencia Artificial en la investigación criminal

Sandra María Piñeiro Vilas
Presidenta de la Audiencia Provincial de Lugo



1. INTRODUCCIÓN: JUSTICIA PENAL E INNOVACIÓN TECNOLÓGICA

En un momento en que la Justicia penal en España sufre un importante déficit de valoración ciudadana, la inteligencia artificial (IA) se presenta como una herramienta prometedora. Pero este refuerzo tecnológico solo será eficaz si se instala con precisión para mejorar la eficiencia y calidad del proceso penal, sin dañar los cimientos éticos que sostienen el sistema de garantías y derechos fundamentales.

Indudablemente la IA puede servir de apoyo a policía, fiscalía y judicatura, pero su implementación debe ser cuidadosa y respetar estrictos criterios normativos, como los que establece el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (en adelante, Reglamento Europeo de IA).

2. USOS DE LA INTELIGENCIA ARTIFICIAL EN EL PROCESO PENAL

Las herramientas de IA aplicadas al proceso penal pueden servir para la predicción y evaluación de riesgos y la investigación de delitos, así como en la generación de contenido probatorio. Pero también a la forma de tramitar las causas penales, dotándolas de mayor celeridad y eficiencia en la gestión de los expedientes judiciales.

Esta última aplicación -menos polémica desde el punto de vista de los derechos fundamentales- puede facilitar la tramitación de causas complejas, aumentar la celeridad del proceso penal y contribuir a una justicia más eficiente y accesible.

No obstante, el uso más sensible y debatido es aquel orientado a la predicción de conductas delictivas, la estimación de riesgos (como la reincidencia o la peligrosidad) o el apoyo a la toma de decisiones judiciales con base en modelos algorítmicos.

3. POTENCIAL PREDICTIVO: ENTRE LA ESTADÍSTICA Y LA DECISIÓN JUDICIAL

Es obvio que los algoritmos no poseen una bola de cristal para predecir el incierto futuro, pero su capacidad para trabajar con grandes volúmenes de datos los convierte en herramientas con un enorme potencial predictivo. Son una suerte de “meteorólogos del comportamiento humano”, que aunque no pueden decirnos con certeza qué ocurrirá, sí pueden señalar, con razonable precisión, dónde hay tormentas en el horizonte. Esta analogía cobra especial relevancia en el ámbito penal, donde la estimación de riesgos —como la reincidencia, la fuga o la comisión de nuevos delitos— no solo es posible, sino deseable, y puede contribuir a la toma de decisiones más informadas, siempre que se actúe con proporcionalidad, rigor metodológico y pleno respeto a los derechos fundamentales.

4. TRADICIÓN CRIMINOLÓGICA: EVALUACIÓN ACTUARIAL DEL RIESGO

La predicción del comportamiento delictivo no es una invención reciente. En criminología, la predicción del comportamiento delictivo ha sido una preocupación constante. La investigación empírica ha buscado modelos que permitan anticipar la probabilidad de que una persona condenada vuelva a delinquir tras su puesta en libertad, lo que ha dado lugar al desarrollo de escalas de riesgo y herramientas de evaluación actuarial.

Estas herramientas combinan variables estáticas (como edad del primer delito o historial delictivo) con variables dinámicas (como consumo de drogas, situación laboral o entorno familiar), a fin de identificar perfiles de riesgo.

En este marco se inscribe el sistema VioGen, utilizado en España para la evaluación de riesgo en casos de violencia de género.

5. VIOGEN: UNA HERRAMIENTA ÚTIL CON LIMITACIONES IMPORTANTES

VioGen ha supuesto un avance indudable en la protección de las víctimas, al proporcionar a las fuerzas y cuerpos de seguridad

del Estado un instrumento homogéneo para valorar el riesgo. Y aunque a menudo se menciona junto a sistemas predictivos modernos, es en realidad un sistema de naturaleza actuarial, ya que funciona con reglas fijas y parámetros predefinidos; se basa en formularios estructurados y cuestionarios que los y las agentes policiales rellenan con datos sobre la víctima, el presunto agresor y las circunstancias del caso, y, a partir de esa información, aplica una fórmula estática (no un algoritmo que aprenda por sí solo) para asignar un nivel de riesgo, revelándose como una herramienta útil para establecer medidas policiales de protección.

La confusión con los sistemas de inteligencia artificial se da porque comparte con la IA cierta apariencia de objetividad automatizada. Pero lo cierto es que, técnicamente, no reúne los requisitos funcionales ni técnicos de un sistema de inteligencia artificial moderno. Y si bien ha supuesto un avance en la protección de víctimas de violencia de género en España, al dotar a las fuerzas policiales de una herramienta estructurada y homogénea para evaluar el riesgo, su uso plantea problemas graves de transparencia, control y respeto a los derechos fundamentales, tanto de las víctimas como de las personas denunciadas, requiriendo no solo una revisión integral mediante su actualización tecnológica, sino la incorporación de elementos de inteligencia artificial explicable, así como transparencia en los criterios y datos utilizados, y un control ju-



**“En criminología,
la predicción del
comportamiento
delictivo ha sido
una preocupación
constante”**

dicial efectivo, sin que quepa la simple incorporación en la motivación de las resoluciones judiciales, y la posibilidad de contradicción. A ello se debe añadir la realización de auditorías externas independientes y la formación especializada de todos los operadores jurídicos y policiales implicados.

6. DE LA EVALUACIÓN ACTUARIAL A LA PREDICCIÓN ALGORÍTMICA: UNA FRONTERA DIFUSA

La evolución hacia modelos algorítmicos más sofisticados ha dado lugar a herramientas que, aunque basadas en parámetros similares a los actuariales, incorporan técnicas de aprendizaje automático (*machine learning*) que permiten identificar patrones complejos de riesgo:

- LSI-R (Level of Service Inventory – Revised), desarrollado por Andrews y Bonta, que es una de las herramientas más utilizadas para evaluar el riesgo de reincidencia y las necesidades criminógenas, el cual analiza factores estáticos (edad del primer delito, antecedentes, etc.) y dinámicos (uso de drogas, relaciones familiares, empleo, etc.).
- STATIC-99, específico para delincuentes sexuales, combina variables fijas como edad, historial de delitos sexuales, número de víctimas, etc., para estimar el riesgo de reincidencia sexual, y es uno de los instrumentos más validados a nivel internacional.
- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), instrumento utilizado en EE. UU., desarrollado por Northpointe, analiza una amplia gama de factores para predecir riesgos de reincidencia y necesidades de supervisión.
- RisCanvi, desarrollado por Departamento de Justicia de Cataluña, evalúa el riesgo de reincidencia y otros indicadores en internos, combinando variables personales, delictivas y conductuales.

Si bien estos sistemas pueden considerarse, con matices, formas de IA predictiva, es fundamental no sobredimensionar sus capacidades, pues no están exentos de limitaciones técnicas, sesgos o riesgos de opacidad. Estos algoritmos de aprendizaje automático funcionan como artesanos de patrones invisibles: buscan relaciones entre variables que a menudo escapan a la lógica humana, hilando con precisión milimétrica las combinaciones de factores que, en el pasado, han coincidido con determinadas conductas delictivas. En el caso de la reincidencia, por ejemplo, los algoritmos analizan un conjunto de datos que puede incluir edad, historial delictivo, tipo de delito cometido, entorno familiar, situación laboral, presencia de adicciones o enfermedades mentales, entre otros. Es como si trazaran un mapa por el que otras personas con trayectorias similares han transitado antes, para calcular qué desvíos o retornos son más probables.

A partir de ahí, elaboran perfiles de riesgo que pueden influir en decisiones tan sensibles como la concesión de un permiso penitenciario, la imposición de una medida cautelar o la evaluación de una posible amenaza para la seguridad pública.

No obstante, su grado de sofisticación, explicabilidad y adaptabilidad no es comparable a la IA avanzada. Lo que, unido al impacto de estos sistemas en derechos

fundamentales, requiere el cumplimiento de estándares éticos, técnicos y jurídicos, especialmente en lo referente a transparencia, no discriminación, control humano y necesidad de auditoría externa.

La predicción algorítmica no se presenta, pues, con la fuerza del sino; es una advertencia estadística.

7. APLICACIONES PRÁCTICAS DE LA IA EN DISTINTAS FASES DEL PROCESO PENAL

Esta tecnología ha demostrado ser útil en diferentes fases del proceso penal:

7.1 Intervención en la fase policial: predicción delictiva y distribución de recursos

En primer lugar, en fase policial, al detectar , como una suerte de radares sociales, dónde puede surgir la próxima infracción, lo que permite una mejor distribución de los recursos policiales.

7.2 Utilidad en el ámbito penitenciario: evaluación del riesgo y toma de decisiones

Pero también en el ámbito penitenciario, evaluando el riesgo de reincidencia violenta o de quebrantamiento de condena, a fin de tomar decisiones sobre el régimen de vida del interno (progresiones en grado, concesión de permisos, etc).

7.3 Incorporación en sede judicial: apoyo a la adopción de medidas judiciales

Pero también en sede judicial, incorporándose como un elemento más para valorar medidas cautelares o condiciones de libertad condicional. Si bien, en este último punto, no cabe desconocer el riesgo de

que las decisiones judiciales dejen de ser fruto de la deliberación razonada para automatizarse y convertirse en el reflejo de una predicción estadística.

8. RIESGOS DE LA JUSTICIA ALGORÍTMICA: SESGOS, DESHUMANIZACIÓN Y PÉRDIDA DE GARANTÍAS

La literatura criminológica ha advertido sobre los límites de esta forma de justicia predictiva. Los peligros de los sesgos algorítmicos están ahí, cuando los datos históricos reflejan desigualdades estructurales —por ejemplo, una sobrerrepresentación de ciertas minorías en las estadísticas delictivas—, pues los modelos aprenden y perpetúan esas distorsiones, incluso aunque no estén expresamente programados para ello. Es decir, los algoritmos no solo heredan el pasado, sino que corren el riesgo de amplificar sus injusticias.

De ahí la importancia de advertir que estos sistemas no deben sustituir el juicio humano, sino complementarlo. Los algoritmos actúan como un microscopio: permiten observar con mayor precisión ciertos patrones y detalles que podrían pasar desapercibidos a simple vista. Pero sigue siendo el juez quien interpreta lo que ve, quien elige el enfoque y quien decide qué conclusiones extraer del análisis. Solo así puede garantizarse que las decisiones se fundamenten no en el automatismo frío de un modelo matemático, sino en la ponderación jurídica y ética del caso concreto.

En definitiva, la IA aplicada a la predicción penal es una herramienta poderosa, pero frágil. Puede contribuir a una justicia más eficiente, pero también puede generar nuevos peligros si se utiliza sin transparencia, sin control humano o sin garantías. La prevención del delito, la gestión penitenciaria y las decisiones judiciales son

territorios complejos, donde la estadística debe convivir con la prudencia y la dignidad de las personas.

9. LA IA COMO ALIADA EN LA INVESTIGACIÓN CRIMINAL Y GENERACIÓN DE PRUEBA

La IA se revela como una poderosa herramienta de investigación, y de generación de prueba o evidencias, apoyando la investigación de delitos y la obtención de medios de prueba.

Son múltiples las herramientas en este punto: desde reconocimiento biométrico (facial, de voz, de emociones, optimización de informes de huellas y ADN, o pericias caligráficas humanas), hasta el procesamiento del lenguaje natural mediante chatbots con aptitud para asistencia a víctimas o agentes encubiertos en investigaciones online.

Los sistemas de análisis de documentos permiten gestionar grandes volúmenes de información. Cabe traer a colación el “caso Rolls Royce”, en el que se utilizó por primera vez en Reino Unido -cuando todavía formaba parte de la UE-, en un expediente judicializado, un sistema de IA de análisis de documentos, que llevó a cabo el análisis de aproximadamente treinta millones de documentos de forma dos mil veces más rápida que un ser humano, con una reducción del coste del 80 %. O el sistema VeriPol, para detectar denuncias falsas de robos con violencia e intimidación presentadas en comisarías, con mejoras notables en coste y tiempo de investigación.

El análisis de malware y contenido ilícito online permite detectar delitos de odio, amenazas o radicalización online, sirviendo como *notitia criminis* pero también como prueba de dichos ilícitos para el enjuiciamiento.

10. IA Y PRUEBA VISUAL: VERIFICACIÓN, ANÁLISIS MASIVO Y TRAZABILIDAD

En un mundo hipervisual y digitalizado, la proliferación de imágenes y vídeos en redes sociales, plataformas de mensajería, cámaras de seguridad y dispositivos personales ha hecho que muchos delitos de extrema gravedad dejen huellas digitales en formato visual. Pero el volumen de datos es inabarcable sin herramientas automatizadas.

Es en este punto donde cabe destacar la IA, como herramienta que permite procesar millones de archivos en tiempos razonables, filtrando y categorizando lo relevante,



“En un mundo hipervisual y digitalizado, la proliferación de imágenes y vídeos en redes sociales, [...] ha hecho que muchos delitos de extrema gravedad dejen huellas digitales en formato visual”

y detectando patrones ocultos o coincidencias que, de otro modo, serían imposibles de encontrar.

Así, en el caso de la pornografía infantil, a través del análisis automático de imágenes y reconocimiento de patrones visuales, podría permitir detectar, clasificar y rastrear material de abuso sexual infantil.

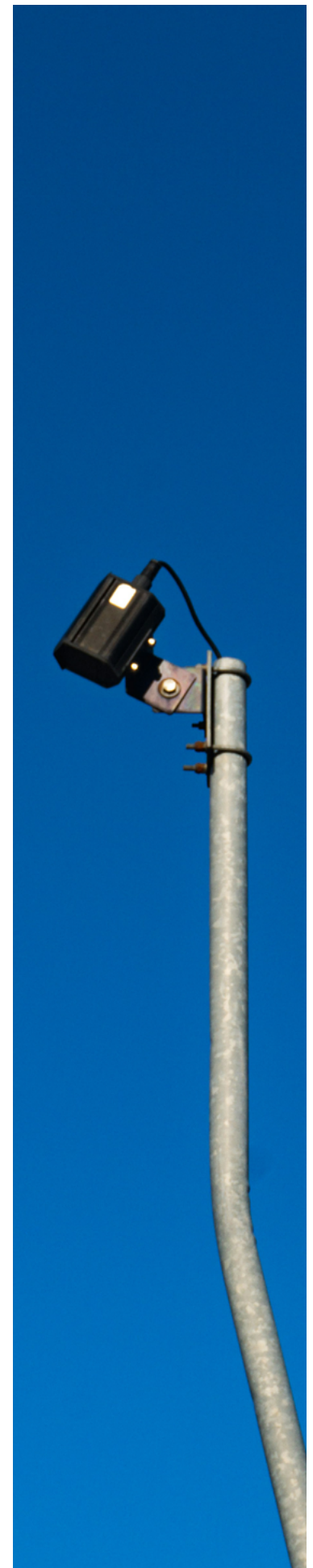
Herramientas como las desarrolladas por Google (API de Content Safety) o Microsoft (PhotoDNA) utilizan IA para generar huellas digitales únicas (hashes) de contenido ilícito, permitiendo identificar automáticamente copias incluso aunque hayan sido alteradas (recortes, filtros, compresión). Esta tecnología también permite desanonimizar a agresores o identificar a víctimas mediante el reconocimiento facial, comparación de escenarios o coincidencias biométricas en múltiples imágenes.

11. APLICACIONES EN EL DERECHO PENAL INTERNACIONAL: VERIFICACIÓN DE CRÍMENES CONTRA LA HUMANIDAD

Por otra parte, en este mundo globalizado, con una miríada de conflictos armados, la IA puede analizar grabaciones de satélite, drones, móviles o redes sociales para identificar violaciones de Derecho Humanitario, tales como ataques a población civil, el empleo de armas prohibidas o el uso de tácticas de limpieza étnica o tortura sistemática.

Proyectos como el de “EyeWitness to Atrocities” (desarrollado por el International Bar Association) o iniciativas de Open Source Intelligence (OSINT) integran IA para verificar la autenticidad, fecha, geolocalización y manipulación de imágenes o vídeos, para su empleo como prueba judicial ante los Tribunales. Estas imágenes pueden ser claves para acreditar patrones sistemáticos de criminalidad, intención genocida o la cadena de mando en la ejecución de crímenes.

De hecho, en la Corte Penal Internacional, la prueba audiovisual analizada por IA ha sido determinante para abrir investigaciones preliminares o emitir incluso órdenes de arresto, como ocurrió en marzo de 2023, cuando emitió órdenes de arresto contra Vladímir Putin y Maria Lvova-Belova por la deportación forzosa de niños ucranianos, o en 2019, cuando abrió una investigación formal y emitió órdenes de arresto por la persecución a la minoría rohingya en Myanmar.



12. GARANTÍAS PROCESALES ANTE LA PRUEBA DIGITAL OBTENIDA MEDIANTE IA

En un mundo saturado de imágenes, la IA permite verificar la autenticidad de vídeos clave (no manipulados, con geolocalización exacta), procesar enormes volúmenes de datos en tiempo útil para los procedimientos judiciales y corroborar patrones de criminalidad sistemática, esenciales en el Derecho penal internacional. Y aunque la decisión judicial sigue en manos humanas, la IA se ha convertido en una aliada técnica imprescindible para transformar imágenes dispersas en pruebas jurídicas siempre que se cumplan los estándares de autenticidad mediante la trazabilidad y verificabilidad de su origen, cadena de custodia, garantizando la integridad del archivo digital desde su obtención hasta su práctica en el plenario, registrando cualquier intervención, garantizando el principio de contradicción, permitiendo a las partes revisar, impugnar y debatir la prueba obtenida o filtrada por IA, lo que conecta con el principio de transparencia, y en todo caso salvaguardando su sometimiento a la valoración judicial humana.

En resumen: toda prueba generada o filtrada mediante IA debe cumplir con las garantías reforzadas de autenticidad, cadena de custodia digital, posibilidad de contradicción y registro de intervenciones, sin renunciar a la valoración judicial humana. Y el uso de IA en la obtención y análisis de prueba debe someterse a un estricto control jurisdiccional y garantizar que las partes puedan debatir y someter a contradicción la fiabilidad de las herramientas utilizadas.

13. DEEPFAKES: UNA AMENAZA EMERGENTE PARA LA VERDAD JUDICIAL

Pero esto conecta también con una nueva problemática asociada al empleo de la IA para la obtención de fuentes de prueba:

los deepfakes, que constituyen una de las amenazas más sofisticadas y disruptivas para el proceso penal contemporáneo. La inteligencia artificial es, al mismo tiempo, el problema y la solución: genera falsificaciones realistas, pero también proporciona herramientas para su detección.

La detección de deepfakes se ha convertido, así, en uno de los nuevos frentes más sensibles para garantizar la integridad de las pruebas en el proceso penal, debido al alto grado de sofisticación técnica y verosimilitud visual y auditiva que estas falsificaciones pueden alcanzar, aparentando total autenticidad al imitar la identidad de una persona real, mediante la suplantación de la voz o del rostro de una persona en un vídeo, o bien simulando declaraciones, amenazas, actos violentos o confesiones que nunca sucedieron, los cuales pueden ser difundidos como noticia o prueba, con finalidades delictivas o desinformativas.

Implicaciones penales y procesales del uso malicioso de deepfakes

El riesgo evidente, en un proceso penal, es que un deepfake puede presentarse como prueba aparentemente inculpativa, pero ser una falsificación perfecta, lo que pone en riesgo la presunción de inocencia, el derecho de defensa y la fiabilidad probatoria. Las implicaciones son evidentes y conllevan la necesidad de un *cambio de paradigma probatorio*.

Tradicionalmente, una grabación audiovisual o sonora tenía valor probatorio, salvo que se demostrara manipulación. Con la aparición de los deepfakes, esa presunción de veracidad ha de invertirse, y jueces y juezas han de extremar las precauciones, adoptando una actitud crítica y cautelosa, especialmente cuando la autenticidad de la prueba sea impugnada. Pero esto también conlleva la necesidad de que las partes -obviamente tanto defensa como acusación- tengan derecho e incluso la obligación para el pleno respeto de los derechos fundamentales en

tensión en el proceso penal y la observancia de las garantías procesales del proceso debido, de solicitar la verificación técnica ante cualquier sospecha, revelándose como absolutamente necesarios los informes periciales tecnológicos, en orden a la obtención de una verdad procesal lo más próxima posible al acontecer histórico objeto del proceso.

Pero, paradójicamente, solo la propia inteligencia artificial puede detectar con eficacia muchos de los deepfakes actuales, ya que el ojo humano -ni siquiera el experto- puede distinguirlos con fiabilidad. Aquí es donde entran en juego los sistemas automatizados de detección de manipulaciones digitales mediante análisis de inconsistencias visuales (parpadeo anormal, sincronización labial imperfecta, sombras incoherentes), examen de metadatos, análisis de audio sintético, que permita detectar patrones de entonación o timbre que no coinciden con grabaciones reales, etc. Ejemplos de estas herramientas ya existentes son p. ej. Microsoft Video Authenticator, Deepware Scanner o Amnesty's Digital Verification Corps, entre otros.

El Reglamento (UE) 2024/1689 de Inteligencia Artificial clasifica los sistemas de IA destinados a evaluar la fiabilidad de pruebas durante investigaciones o juicios penales como sistemas de "alto riesgo" (Anexo III, sección 6.c), lo que conlleva que su uso está sujeto a requisitos estrictos de transparencia, supervisión humana y rendición de cuentas. Deben garantizar explicabilidad técnica, protección de datos (LO 7/2021 y RGPD), y no generar decisiones automatizadas sin control judicial. Requieren auditorías, evaluación de impacto y control público.

El uso malicioso de deepfakes plantea no solo un problema probatorio, sino un riesgo penal autónomo, especialmente cuando se generan para falsificar pruebas o inducir a error a un tribunal, se utilizan para dañar la imagen pública o personal de una víctima, testigo o juez o se difunden masivamente como parte de campañas de desinformación procesal.

La cuestión plantea además otras reflexiones acerca de si contamos actualmente con las herramientas para perseguir estas conductas con los actuales tipos penales, o si urge la tipificación penal autónoma de la creación y difusión maliciosa de deepfakes con finalidad probatoria falsa o calumniosa, la introducción en las leyes procesales de prohibiciones específicas en el ámbito procesal penal sobre el uso de pruebas digitales no verificadas, o incluso la creación de un sistema forense estatal o incluso europeo especializado en detección y certificación de autenticidad de pruebas digitales para conjurar los expresados riesgos.

14. LA TRANSFORMACIÓN DIGITAL DEL PROCESO PENAL MEDIANTE INTELIGENCIA ARTIFICIAL

Aparte de las expresadas utilidades en la investigación penal, la incorporación de la IA al proceso penal y a la gestión de expedientes judiciales permitiría una transformación estructural del modelo de justicia.

A través de sistemas inteligentes, transparentes y supervisados, es posible liberar recursos humanos, reducir tiempos de espera, mejorar la calidad de las resoluciones y garantizar un mejor acceso a la tutela judicial efectiva.

Obviamente, cada avance debe ir acompañado de un marco normativo garantista, de protocolos éticos y de una formación continua del personal judicial, para evitar una tecnificación acrítica que ponga en riesgo los derechos fundamentales.

14.1 Una oportunidad estructural para modernizar la Justicia: del expediente digital a la tramitación inteligente

Más allá de mejorar la eficiencia y calidad del proceso penal, y de la digitalización de

contenidos, es necesario transformar las prácticas procesales y automatizar tareas mecánicas mediante herramientas de IA, garantizando al mismo tiempo la celeridad, la precisión y la protección de derechos.

14.1.1 Automatización de tareas rutinarias y de descarga de trabajo

Así, desde diversos sectores se propugna la automatización de tareas rutinarias y de descarga de trabajo, con el objeto de reducir carga de trabajo en los juzgados de guardia y mejorar el control. Y así se señala la posibilidad de la gestión de comparecencias apud acta, hasta la textualización automática de declaraciones grabadas en sede judicial en tiempo real, conservando la literalidad y ahorrando tiempo a judicatura y funcionariado, al tiempo que facilitaría la inclusión directa en resoluciones judiciales; también se señala la posibilidad de la IA aplicada a traducciones de documentos y declaraciones, lo cual exige garantizar alta fiabilidad y permitir la impugnación por las partes así como la revisión por parte de operadores humanos. Incluso el análisis automatizado de escritos procesales, a fin de analizar denuncias y querellas, verificar requisitos formales (poder, tasa, etc), determinar jurisdicción y competencia o para generar alertas, liberando a los operadores jurídicos de tareas mecánicas.

14.1.2 Clasificación, priorización y alerta automatizada de asuntos

Pero dentro de esa aspiración de mejora en eficiencia en la tramitación de los expedientes judiciales, cabe plantearse otras opciones, como la clasificación y priorización automatizada de asuntos, por temas, detectando urgencias (riesgo para menores, violencia de género, prescripción próxima) o asignando prioridades y bandejas de entrada según criterios objetivos (plazo, complejidad, impacto), lo que permitiría una gestión proactiva y racionalizada, especialmente útil en órganos con colapso o alta litigiosidad.

14.1.3 IA generativa para la redacción asistida de resoluciones

Y, dado que el cuello de botella se sitúa en la resolución de casos, surge la posibilidad de redacción asistida de resoluciones y documentos procesales. Así, los sistemas de IA generativa podrían sugerir borradores de providencias o autos simples (incoación, admisión, archivo), proponer extractos normativos o jurisprudenciales vinculados al tipo de asunto e incluir automáticamente datos del expediente (partes, fechas, hechos relevantes).

14.1.4 Apoyo a la labor jurisdiccional mediante análisis contextualizado

La IA podría emplearse con aquellos objetivos para detectar en un escrito las cuestiones jurídicas clave, relacionarlas con jurisprudencia aplicable actualizada (TS, TJUE, TEDH), e incluso identificar doctrina contradictoria o sentencias relevantes del propio órgano judicial, con el consecuente ahorro de horas de búsqueda a jueces, juezas y letrados y letradas de la Administración de Justicia, promoviendo mayor coherencia y calidad argumentativa.

14.1.5 Gestión inteligente del calendario judicial

Y ya en un plano más organizativo, no cabe desconocer la eventual utilización de la IA para una gestión inteligente del calendario judicial, por cuanto puede cruzar la duración media de vistas, la agenda de partes intervinientes o la disponibilidad de Salas, y, en consecuencia, proponer -bajo supervisión humana- fechas razonables para celebrar vistas o audiencias, evitando suspensiones innecesarias. Tales intervenciones se revela-

rían como particularmente útiles en partidos judiciales con alta rotación o jurisdicciones con agenda colapsada.

14.1.6 Predicción de carga de trabajo y análisis de rendimiento

E incluso, cabría valorar la virtualidad de la IA para la predicción de carga de trabajo y análisis de rendimiento, por cuanto aplicando IA a los metadatos judiciales se podrían generar mapas de saturación por órgano, sección o jurisdicción, alertas de congestión futura e incluso recomendaciones para redistribución de personal o refuerzo temporal, lo que facilitaría la toma de decisiones estratégicas por parte del CGPJ o las Consejerías de Justicia.

14.2 IA como garantía del cumplimiento normativo en protección de datos

LA IA también se presenta como una herramienta para garantizar la LO 7/2021 y RGPD, por cuanto puede permitir la detección automática de datos personales o sensibles en documentos que van a ser publicados, la anonimización automática de resoluciones para protección de menores, víctimas o testigos, e incluso la detección de documentos que incumplen requisitos de seguridad digital (malwares, enlaces sospechosos, manipulaciones).

14.3 Hacia una justicia más accesible e inclusiva

Otro aspecto a valorar, dentro de la aspiración de aproximar la Administración de justicia a la ciudadanía es la posible traducción automatizada combinada con simplificación del lenguaje jurídico para garantizar accesibilidad y la adaptación de textos procesales a lenguaje claro, ajustado a personas con discapacidad cognitiva, migrantes o menores. Esto contribuiría a una justicia más inclusiva, sin barreras lingüísticas ni técnicas.

15. CONSIDERACIONES FINALES: UNA OPORTUNIDAD TRANSFORMADORA BAJO CONDICIONES DE CAUTELA Y GARANTÍA

En definitiva, es cierto que el uso de la IA en la investigación criminal, en la obtención de medios y fuentes de prueba, así como en la tramitación de los expedientes judiciales, abre muchas posibilidades, pero también plantea importantes desafíos:

En primer lugar, por la línea roja de las garantías procesales y derechos fundamentales, lo que no solo requiere transparencia y explicabilidad sino supervisión judicial activa, justificando expresamente el uso o rechazo de sistemas de IA y entendiendo sus implicaciones técnicas y jurídicas.

Por otra parte, los algoritmos pueden reproducir y amplificar sesgos existentes en los datos, por lo que la formación y evaluación ética de estos sistemas se vuelve esencial para evitar desigualdades.

Se ha demostrado también que el desuso o falta de mantenimiento de los algoritmos puede poner en riesgo la calidad de las decisiones adoptadas.

También ha de conjurarse el riesgo de desigualdad tecnológica, lo que puede agravar las desigualdades en la calidad de la justicia.



Por otra parte, muchos sistemas tratan datos sensibles (biometría, salud, origen étnico), lo que requiere un marco de garantías robusto, que limite su uso a casos estrictamente necesarios y bajo autorización expresa.

Resulta indeclinable la formación de los operadores jurídicos: fiscales, jueces y juezas, policías y abogados y abogadas deben estar preparados para entender el funcionamiento, los límites y las posibilidades de la IA. De lo contrario, el uso se vuelve arbitrario o ciego.

En todo caso, España deberá adaptar su legislación penal y procesal para dar cabida al uso de estos sistemas, siguiendo los márgenes permitidos por el Reglamento Europeo de IA, no solo mediante la reforma de las Leyes de Enjuiciamiento, sino mediante la creación de protocolos de control sobre los sistemas utilizados y el establecimiento de registros de sistemas de alto riesgo.

Y es que la IA representa una herramienta de gran valor para mejorar la eficiencia del proceso penal, pero su implementación debe estar guiada por principios de cautela, legalidad, proporcionalidad y respeto a los derechos fundamentales. El Reglamento Europeo de IA proporciona un marco sólido, pero será responsabilidad de cada Estado, y, en particular, del poder legislativo y del poder judicial, asegurar su correcta aplicación.

La clave no es prohibir el uso de la IA en justicia penal, sino garantizar su utilización ética, responsable y justa.

La máquina no tiene alma ni conciencia: es el espejo de quien la forja. La bondad o la ruina no nacen de su engranaje, sino del propósito que la mueve. Pues la ética habita en la mano que programa, en la mirada que decide; somos nosotros quienes le damos alma.

Bibliografía

- Alemán Aróstegui, L. (2023). El uso del RisCanvi en la toma de decisiones penitenciarias. *Estudios Penales y Criminológicos*, 44. <https://doi.org/10.15304/epc.44.8884>
- Andrés Pueyo, A. (2016). ¿Es técnicamente posible anticipar la reincidencia delictiva? El protocolo RisCanvi en las prisiones de Cataluña. En IX Jornadas de ATIP Almagro (pp. 55–78).
- Borges Blázquez, R. (2024). Algoritmización de la concesión de medidas cautelares en el proceso penal para la protección de víctimas de violencia de género. ¿Es capaz VIOGÉN de interpretar el periculum in mora? *Actualidad Jurídica Iberoamericana*, 384–407.
- Cuatrecasas Monforte, C. (2024) La inteligencia artificial en el proceso penal. En *Inteligencia artificial y justicia, Formación a Distancia 2 - 2024*. <https://doi.org/10.62659/FA2400208>
- Cueto Santa Eugencia, E. (2025). Herramientas de predicción criminal y su impacto en nuestro proceso penal. Universidad Pontificia Comillas (ICADE). <https://doi.org/10.12795/IESTS-CIENTIA.2025.i01.01>
- Dribia Data & Direcció General d'Afers Penitenciaris. (2024). Informe Tiresias: Auditoria de l' algorisme RisCanvi (Versión del 9 de enero de 2024). Generalitat de Catalunya.
- Estévez Mendoza, L. (2021). Inteligencia artificial y violencia contra las mujeres: ¿Funcionan los sistemas automatizados de evaluación del riesgo? *Perspectivas*. *Revista de Ciencias Jurídicas y Políticas*, 127–141.
- García-Varela Iglesias, R. (2024). *E-Justicia inteligente: navegando por los desafíos de la arquitectura de la información y la IA*. A Coruña: Editorial Colex.
- González-Álvarez, J. L., Santos-Hermoso, J. & Camacho-Collados (2020). Policía predictiva en España. Aplicación y retos de futuro. *Behavior & Law Journal*, 26–41. <https://doi.org/10.47442/blj.v6.i1.75>
- Gordo, Á., y Rubio-Martín, M. J. (2024). Incertidumbres algorítmicas en torno a las violencias de género. *Revista Española de Sociología*. <https://doi.org/10.22325/fes/res.2024.225>
- Pereira Puigvert, S. (2024). La justicia y el proceso penal en clave algorítmica: Nuevos enfoques, nuevos riesgos. *Ius et Scientia*, 66–79. <https://doi.org/10.12795/IESTS-CIENTIA.2024.i02.03>
- Pérez Salazar, B. (2024). La criminología predictiva: ¿un futuro próximo o una ficción en lontananza? *Novum Jus*, 343–369. <https://doi.org/10.14718/NovumJus.2024.18.3.13>
- Roa Avella, M. P., Sanabria-Moyano, J. E., Dinas-Hurtado, K. (2020). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. Universidad Militar Nueva Granada, Bogotá.
- Rueda Soriano, Y., Navarro Blasco, E. (2019). Los sistemas actuariales de prevención y gestión de riesgos en el ámbito penitenciario. *Juezas y jueces para la democracia*, núm. 94, pp. 59–68.
- Tonry, M. (2019). Predictions of dangerousness in sentencing: Déjà vu all over again. En *American sentencing: What happens and why?* (pp. 439–482). The University of Chicago Press. <https://doi.org/10.1086/701895>

It's a Blast!



Fallout

Fallout
(Temporada 1)

Diego Gutiérrez Alonso
Magistrado



Plataforma:
Prime Video.

Valoraciones:

- **IMDb:**
8.3/10
basado en más
de 330.000 votos
- **Rotten Tomatoes:**
93%
de crítica,
93%
de audiencia
- **FilmAffinity:**
6.2/10
basado en más
de 15.500 votos



La primera temporada de Fallout, adaptación televisiva del célebre videojuego de Bethesda, se presenta como un producto de notable factura visual y narrativa que logra trasladar con fidelidad la atmósfera del yermo postnuclear. Más allá de su valor como serie de aventuras distópicas, Fallout abre un campo fértil para reflexionar sobre la relación entre tecnología, capitalismo y humanidad en clave contemporánea.

En el universo de Fallout, la promesa tecnológica se ha convertido en trampa. La obsesión por el progreso, encarnada en robots domésticos, armamento automatizado y gigantescos ordenadores de control, conduce a un mundo devastado por el uso irresponsable de la ciencia. Aunque la inteligencia artificial no es un tema central de la trama, el trasfondo invita a establecer paralelismos con los debates actuales: sistemas opacos que deciden sobre la vida de las personas, la confianza ciega en soluciones “inteligentes” y la difusa frontera entre progreso y deshumanización. Los Vaults —supuestos refugios de salvación convertidos en laboratorios de experimentación social— son la metáfora perfecta del riesgo de delegar en la tecnología, sin supervisión ética, la gestión del destino colectivo.

La crítica al capitalismo es otro eje ineludible. Vault-Tec, la corporación que diseña y controla los refugios, representa el paradigma de un sistema donde el beneficio económico prima sobre la vida humana. La serie despliega con acierto la ironía de un capitalismo retrofuturista que vende seguridad mientras condena a la humanidad a la catástrofe. Desde esa óptica, Fallout resuena con discusiones actuales sobre el poder de las grandes tecnológicas y su capacidad de moldear realidades sociales al margen de los intereses ciudadanos.

La adaptación televisiva también merece reconocimiento por su estrecho vínculo con los videojuegos originales. La estética de los escenarios, el humor negro, la crudeza de la violencia y los dilemas morales trasladan al espectador la esencia del juego, sin caer en el fan service superficial. La narrativa mantiene el mismo trasfondo filosófico: en un mundo atravesado por mutantes, máquinas y comunidades rotas, la pregunta decisiva sigue siendo qué significa ser humano cuando lo tecnológico y lo biológico se entremezclan sin fronteras claras.

En definitiva, Fallout es más que una serie de aventuras postapocalípticas: es un recordatorio de cómo el progreso tecnológico, cuando se somete sin freno a la lógica del capital, puede convertirse en una amenaza mayor que las bombas nucleares. En un momento en que la inteligencia artificial despierta temores y esperanzas, Fallout nos devuelve una advertencia incómoda: el futuro no depende solo de la potencia de nuestras máquinas, sino de la ética con la que decidamos utilizarlas.

