



TRABAJO FIN DE GRADO

GRADO EN DERECHO

CURSO ACADÉMICO 2024 / 2025

TÍTULO:

**“ANÁLISIS DE SISTEMAS ALGORÍTMICOS DE
EVALUACIÓN DE RIESGOS DE REINCIDENCIA CRIMINAL”**

WORK TITLE:

**“ANALYSIS OF ALGORITHMIC RISK ASSESSMENT
SYSTEMS FOR CRIMINAL RECIDIVISM”**

AUTOR: NICOLÁS PÉREZ OCHOA

DIRECTOR: JOSÉ IGNACIO SOLAR CAYÓN

Santander, a 30 de Agosto de 2025

ÍNDICE

1. INTRODUCCIÓN	3
2. MARCO CONCEPTUAL: DEFINICIÓN DE INTELIGENCIA ARTIFICIAL EN EL CONTEXTO JURÍDICO	6
2.1. ¿Qué es la Inteligencia Artificial?	6
2.2. Definiciones de Inteligencia Artificial en el ámbito jurídico	8
3. USOS DE LA INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN DE JUSTICIA	12
3.1. Clasificación de tareas dentro de la Administración de Justicia	12
3.2. Principales ventajas y riesgos	14
4. SISTEMAS ALGORÍTMICOS DE EVALUACIÓN DE RIESGOS DE REINCIDENCIA CRIMINAL	17
4.1. Funcionamiento básico	19
4.2. Principales modelos existentes: COMPAS en EE.UU.	20
5. PRINCIPALES PROBLEMAS QUE PLANTEA SU USO	22
5.1. Dificultades de transparencia	23
5.2. Imparcialidad y sesgos algorítmicos	24
5.3. Uso de “perfiles” para decisiones individualizadas: impacto en la presunción de inocencia	25
6. COMPATIBILIDAD CON EL MARCO JURÍDICO EUROPEO	26
7. CONCLUSIONES	28
8. BIBLIOGRAFÍA	32

RESUMEN

Este trabajo estudia los sistemas algorítmicos de evaluación del riesgo de reincidencia criminal, con especial atención al modelo COMPAS en Estados Unidos. Se analizan sus ventajas y riesgos, en particular los problemas de transparencia, sesgos y compatibilidad con principios como la presunción de inocencia y la igualdad ante la ley. Asimismo, se examina su encaje en el marco jurídico europeo, que los considera de alto riesgo y exige estrictos requisitos de control y transparencia. El trabajo concluye destacando la necesidad de garantizar un uso responsable de estas herramientas que respete los derechos fundamentales.

Palabras clave: Inteligencia Artificial, Administración de Justicia, Algoritmos, COMPAS, Reincidencia, Derechos Fundamentales

ABSTRACT

This paper examines algorithmic systems for assessing the risk of criminal recidivism, focusing on the COMPAS model in the United States. It analyzes their advantages and risks, particularly issues of transparency, bias, and compatibility with principles such as presumption of innocence and equality before the law. The study also reviews their position within the European legal framework, which classifies them as high-risk and requires strict control and transparency measures. It concludes by stressing the need to ensure a responsible use of these tools that safeguards fundamental rights.

Key words: Artificial Intelligence, Administration of Justice, Algorithms, COMPAS, Recidivism, Fundamental Right

1. INTRODUCCIÓN

En las últimas décadas, la inteligencia artificial (IA) ha dejado de ser una idea reservada a la ciencia ficción para convertirse en uno de los fenómenos más influyentes y transformadores del siglo XXI. Desde sus primeras formulaciones teóricas en la mitad del siglo XX, la IA ha evolucionado hasta convertirse en un campo interdisciplinario que combina la informática, la estadística, la neurociencia, la filosofía y otras ramas del conocimiento con el objetivo de crear sistemas capaces de emular habilidades cognitivas humanas como el aprendizaje, el razonamiento, la percepción y la toma de decisiones.

Actualmente, la IA está presente en múltiples aspectos de la vida cotidiana: desde los aparentemente sencillos algoritmos que personalizan contenidos en redes sociales y plataformas de streaming o los que nos ofrecen publicidad personalizada que se ajuste a nuestros gustos e intereses; hasta las complejas estructuras de software que están detrás de los asistentes virtuales, los vehículos autónomos, los diagnósticos médicos asistidos por ordenador o los sistemas de reconocimiento facial.

Esta irrupción de la inteligencia artificial en nuestro día a día ha transformado profundamente diversos aspectos de la vida social, económica y administrativa. La Administración de Justicia no ha permanecido ajena a este fenómeno, incorporando progresivamente sistemas digitales y herramientas algorítmicas con el propósito de agilizar los procedimientos, optimizar recursos públicos y mejorar la toma de decisiones.

De esta manera, el mundo del derecho, con su carácter formal y repleto de solemnidades, se ha topado de manera inevitable con la denominada justicia digital, confluyendo en un escenario en constante evolución, en el que la tecnología convive con principios fundamentales del Derecho que deben garantizarse en todo momento, como el debido proceso, la tutela judicial efectiva o la igualdad ante la ley.

Entre las aplicaciones más controvertidas se encuentran los sistemas algorítmicos de evaluación de riesgos de reincidencia criminal. Estos programas, basados en técnicas de machine learning y aprendizaje automático, analizan variables personales, sociales y delictivas con el objetivo de predecir la probabilidad de que un individuo vuelva a delinquir. Uno de los ejemplos más conocidos es el sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), utilizado en distintos estados de Estados Unidos para orientar decisiones judiciales relativas a la fijación de medidas cautelares o la concesión de la libertad condicional.

El empleo de este tipo de herramientas plantea numerosos interrogantes jurídicos y éticos. Por un lado, estos sistemas se presentan como una solución capaz de dotar al sistema de una mayor objetividad y eficiencia en la administración de justicia, reduciendo la discrecionalidad judicial y contribuyendo a una mejor gestión del sistema penitenciario. Por otro lado, han sido numerosas las críticas que señalan a estos sistemas de peligrosos debido a distintos factores tales como: la opacidad de los algoritmos, la posible reproducción de sesgos discriminatorios, la dificultad de garantizar la transparencia, así como la tensión que generan con principios básicos como la presunción de inocencia o la individualización de la pena.

En cualquier caso, el hecho de que un sistema como COMPAS haya generado tantas polémicas pone de manifiesto la necesidad de llevar a cabo un estudio pormenorizado y exhaustivo acerca de su funcionamiento, sus limitaciones, los posibles riesgos y consecuencias de su uso, así como las condiciones de su implementación. Como veremos, más allá de sus promesas y logros, la inteligencia artificial plantea desafíos de gran relevancia que no deben ser pasados por alto. Cuestiones tan fundamentales como la privacidad de los datos, la discriminación algorítmica, la transparencia de los sistemas, la dependencia tecnológica y el futuro del empleo humano en un mundo cada vez más automatizado han hecho que el uso de la IA se haya convertido en un intenso objeto de debate tanto académico como público.

En este trabajo se abordará, en primer lugar, el marco general de la inteligencia artificial en la Administración de Justicia, para posteriormente centrarse en el análisis de los sistemas algorítmicos de evaluación del riesgo de reincidencia criminal, con especial atención al sistema COMPAS en Estados Unidos. Se valorarán tanto los aspectos positivos como negativos de estos sistemas, teniendo en cuenta las críticas recibidas.

El objetivo principal del presente Trabajo de Fin de Grado es analizar el funcionamiento y las implicaciones jurídicas de estos sistemas. De manera complementaria, se pretende reflexionar sobre la viabilidad de su implantación en otros sistemas jurídicos y sobre los límites que deberían establecerse para garantizar el respeto a los derechos fundamentales.

2. MARCO CONCEPTUAL: DEFINICIÓN DE INTELIGENCIA ARTIFICIAL EN EL CONTEXTO JURÍDICO

2.1. ¿Qué es la Inteligencia Artificial?

Considero un buen punto de partida para adentrarnos en esta temática el delimitar el significado de inteligencia artificial. Un vistazo rápido a la definición dada por un diccionario nos bastará para darnos cuenta de que resulta fútil e insuficiente¹.

Sin embargo, a la hora de buscar una definición más completa, veremos que no se trata de una tarea en absoluto sencilla puesto que debemos atender a multitud de elementos y particularidades. La ambigüedad, la falta de consenso, la mezcla de aspectos técnicos y filosóficos, y sobre todo, la continua evolución de esta disciplina, son factores clave que dificultan establecer una definición clara y precisa de este concepto, generando confusión sobre qué debe incluirse dentro de la categoría de inteligencia artificial y qué no.

En el debate sobre la definición de inteligencia artificial, resulta de interés recordar que, históricamente, la disciplina no ha seguido una única línea de investigación, sino que ha oscilado entre diferentes concepciones de lo que entendemos por “inteligencia”.

Tal como señalan RUSSELL y NORVIG², algunos investigadores han entendido la IA como la capacidad de una máquina de imitar o reproducir fielmente un comportamiento humano; mientras que otros han optado por un enfoque vinculado a la racionalidad, es decir, la capacidad de una máquina de “hacer lo correcto” en un contexto determinado. A ello se añade una segunda distinción: entre quienes conciben la inteligencia como un proceso interno de razonamiento,

¹ La Real Academia Española define la inteligencia artificial de la siguiente manera: “f. Inform. Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico”.

² RUSSELL, S., & NORVIG, P., *Artificial Intelligence: A Modern Approach*, 4th Global Ed., Pearson Global Editions, 2021, p. 19.

y quienes la definen atendiendo a su manifestación externa, es decir, al comportamiento observable.

Un ejemplo paradigmático de lo que entendemos por “actuar humanamente” es el llamado “test de Turing”, propuesto por Alan Turing en 1950, el cual parte de la siguiente pregunta como premisa: “¿pueden pensar las máquinas?”. Turing se propone sustituir esta pregunta filosófica por un criterio empírico y objetivo: si un interrogador humano no pudiera distinguir, a través de preguntas y respuestas escritas, si dialoga con una persona o con una máquina, podría afirmarse que esta última demuestra un comportamiento inteligente. Para superar dicho test, un sistema debería incorporar múltiples capacidades: procesamiento del lenguaje natural, representación del conocimiento, razonamiento automatizado y aprendizaje automático³.

En cualquier caso, en la actualidad, la mayor parte de la comunidad investigadora ha apartado su mirada del test de Turing, para centrar sus esfuerzos en el estudio de los principios generales de la inteligencia, los cuales van mucho más allá de la mera imitación del comportamiento humano. RUSSEL y NORVIG comparan este planteamiento con el éxito de la aviación humana, la cual tuvo lugar gracias a que los ingenieros estudiaron y aplicaron los principios y leyes de la aerodinámica; no gracias a que tratasen de crear máquinas que imitasen el comportamiento de las aves⁴.

Este trasfondo histórico y conceptual pone de relieve que la IA puede definirse desde múltiples perspectivas, y explica por qué, en el ámbito jurídico, resulta difícil alcanzar una definición unívoca. Precisamente por ello, los textos normativos, sobre los que hablaremos más adelante optan por definiciones funcionales y operativas, orientadas a establecer límites regulatorios más que a resolver debates filosóficos o técnicos.

³ TURING, A. M., “Computing Machinery and Intelligence”, *Mind*, volume LIX, Issue 236, October 1950, pp. 433-460, (p. 433) DOI: <https://doi.org/10.1093/mind/lix.236.433>.

⁴ RUSSELL, S., & NORVIG, P., *op cit*, p. 20.

2.2. Definiciones de Inteligencia Artificial en el ámbito jurídico

A pesar de ser la inteligencia artificial un tema que ha suscitado una atención creciente en los últimos tiempos, la realidad es que aún no existe una definición única y universalmente aceptada sobre el término⁵. Dada la dificultad que presenta delimitar de manera concreta y exacta lo que entendemos por inteligencia artificial, debido a la diversidad de perspectivas que pueden adoptarse, puede resultar enriquecedor analizar cómo distintas organizaciones internacionales han definido este concepto en los últimos años.

Un documento que resulta de gran importancia en la materia que nos atañe es la *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, la cual fue adoptada en 2018 por la Comisión Europea para la Eficiencia de la Justicia (CEPEJ) del Consejo de Europa. Dicho documento consta de un glosario donde recoge, entre otras, la definición de inteligencia artificial como “un conjunto de métodos, teorías y técnicas científicas cuyo objetivo es reproducir, mediante una máquina, las capacidades cognitivas de los seres humanos. Los desarrollos actuales buscan que las máquinas puedan realizar tareas complejas que anteriormente eran llevadas a cabo por humanos (*A set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings. Current developments seek to have machines perform complex tasks previously carried out by humans*)”⁶.

Bajo mi punto de vista, esta definición acierta a la hora de poner énfasis en que las máquinas dotadas de inteligencia artificial podrían asumir tareas complejas que hasta ahora habían estado reservadas al ser humano. Este matiz resulta fundamental, ya que permite dimensionar el carácter disruptivo de la IA no solo como una herramienta de apoyo, sino también como un potencial sustituto en ámbitos donde tradicionalmente se consideraba imprescindible la intervención

⁵ “*Qué es la Inteligencia Artificial*”, Plan de Recuperación, Transformación y Resiliencia, 19 de abril de 2023. Disponible en: <https://planderrecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr> (consultado el 12 de julio de 2025).

⁶ EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, Council of Europe CEPEJ (2018). p. 69 (glossary).

humana. Esta idea de máquinas sustituyendo a humanos permite comprender de forma inmediata la dimensión práctica y transformadora de la IA.

Sin embargo, considero que esta definición se encuentra lejos de conseguir el objetivo de ofrecer una descripción integral acerca de la inteligencia artificial, pues presenta determinadas limitaciones que conviene señalar.

En primer lugar, se trata de una definición amplia y algo abstracta, lo que puede generar dificultades prácticas a la hora de determinar con precisión qué sistemas deben considerarse inteligencia artificial, particularmente en el marco judicial.

En segundo lugar, al centrarse en la idea de reproducir habilidades humanas, corre el riesgo de transmitir una visión demasiado simple y reduccionista, sin tener en cuenta que muchos sistemas de IA actuales no buscan imitar al ser humano, sino optimizar procesos a través de métodos estadísticos y de aprendizaje automático que no necesariamente replican las capacidades de cognición humana.

Como explica CRIADO, J.I., basándose en la obra de RUSSEL y NORVIG, podemos hablar de “máquinas desarrollando tareas mediante comportamientos inteligentes que reproducen los humanos” y por otro lado de “máquinas comportándose racionalmente percibiendo lo que sucede en el ambiente y tomando decisiones para alcanzar objetivos concretos”. En otras palabras, el desarrollo de estas tecnologías no trata únicamente de emular el comportamiento humano sino que se centra en crear máquinas autónomas capaces de actuar correctamente de acuerdo con la situación en la que se encuentren. No se busca simplemente una mera repetición de lo que tal vez haría un humano sino que cada vez “se comienza a dar una mayor importancia a la cuestión de capacidades que desarrolla la IA”⁷.

En consecuencia, si bien la definición aportada por el Consejo de Europa deja evidencia de la capacidad transformadora de la IA, y su potencial para alterar

⁷ CRIADO, J.I., “Inteligencia Artificial (y Administración Pública)”, *Eunomia. Revista en Cultura de la Legalidad*, 20, 2021, pp. 348-372, (pp. 351-352). DOI: <https://doi.org/10.20318/eunomia.2021.6097>.

estructuras laborales, sociales y jurídicas, resulta insuficiente para abordar de manera integral las posibilidades de la IA de actuar racionalmente y de manera acorde a la situación que se trate.

Por otro lado, la OCDE, define sistema de IA como “un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos de entrada que recibe, cómo generar información de salida como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos reales o virtuales. Una vez implementados, los distintos sistemas de IA presentan diversos niveles de autonomía y varían en su capacidad de adaptación (*An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment*)”⁸.

La definición propuesta por la OCDE tiene como punto fuerte su claridad y operatividad. Explica de manera directa cómo funcionan los sistemas de inteligencia artificial: percibiendo una información, la cual es procesada, para generar una respuesta, es decir, una salida, la cual puede tener distintas formas como “predicciones, contenidos, recomendaciones o decisiones”.

Asimismo, introduce dos elementos clave de la IA contemporánea, que no encontramos en la definición propuesta por el Consejo de Europa. Estos elementos son la autonomía y la capacidad de adaptación, los cuales constituyen precisamente los rasgos que diferencian a estos sistemas de otras tecnologías automatizadas más simples⁹.

Por autonomía, podemos entender que implica que una vez desplegados, los sistemas de IA pueden operar sin supervisión constante del ser humano,

⁸ OCDE. (2020). *Legal recommendation 0449*, 22 de mayo de 2019. Recuperado de: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>].

⁹ PRESNO LINERA, M.A, & MEUWESE, A., “La regulación de la inteligencia artificial en Europa”, *Teoría y Realidad Constitucional*, (54), 2024, pp. 131–161 (p. 142) DOI: <https://doi.org/10.5944/trc.54.2024.43310>

ejecutando tareas complejas de forma independiente dentro de los parámetros definidos. Por su parte, cuando hablamos de capacidad de adaptación hacemos referencia a la facultad de las máquinas de modificar su comportamiento a partir de nuevas entradas de datos o de la experiencia acumulada, lo que les confiere un dinamismo ausente en los sistemas automatizados convencionales.

En este sentido, la definición de la OCDE resulta funcional porque aporta un criterio claro de diferenciación, evitando la sobre inclusión dentro de lo que entendemos por IA, de herramientas que no presentan ningún rasgo de autonomía en absoluto, o las que no tengan ningún tipo de capacidad de adaptación.

Por su parte, el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 define en su artículo 3 el significado del término “sistema de IA” como “un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales”¹⁰.

Como podemos comprobar, esta última definición es bastante similar a la ofrecida por la OCDE, haciendo también referencia a cómo los sistemas de IA pueden influir “en entornos físicos o virtuales”. Este añadido busca abarcar tanto los sistemas de IA que tienen un impacto material directo en la realidad física como aquellos cuyo campo de acción se limita a lo digital o informacional. El término “entornos físicos o virtuales” funciona como un concepto amplio y neutral, pensado para no dejar fuera ninguna aplicación relevante de la inteligencia artificial¹¹.

Por lo demás, puede afirmarse que esta definición resulta adecuada en términos generales. Si bien en una primera lectura podría parecer excesivamente vaga o

¹⁰ «DOUE» núm. 1689, de 12 de julio de 2024

¹¹ PRESNO LINERA, M.A. & MEUWESE, A., *op cit.*, p. 141.

ambigua, en realidad esa amplitud constituye una de sus principales virtudes. Definir la inteligencia artificial con un margen de flexibilidad permite que la norma no quede rápidamente obsoleta frente a la constante evolución tecnológica, y que los futuros desarrollos en este ámbito se encuentren igualmente bajo el paraguas regulador. De este modo, se evita el riesgo de que un marco normativo demasiado rígido o cerrado termine siendo ineficaz al no contemplar innovaciones que, aunque no encajen en la concepción técnica actual de la IA, generen impactos semejantes o incluso mayores en la sociedad¹².

En definitiva, delimitar con exactitud el significado de IA no es algo sencillo y aunque la aparente indeterminación de la definición pueda en principio verse como una carencia, debemos entender que se trata en realidad de una estrategia consciente de técnica legislativa orientada a garantizar su eficacia en el tiempo. En un contexto marcado por la rápida evolución de la inteligencia artificial, resulta imprescindible que la regulación encuentre un equilibrio entre la seguridad jurídica y la necesaria flexibilidad. Solo de este modo es posible ofrecer un marco normativo que, sin perder claridad, sea lo suficientemente amplio como para abarcar tanto los sistemas actualmente existentes como aquellos que puedan desarrollarse en el futuro.

3. USOS DE LA INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN DE JUSTICIA

3.1. Clasificación de tareas dentro de la Administración de Justicia

Cuando se discute sobre la incorporación de la inteligencia artificial en la Administración de Justicia, es habitual que la primera imagen que acuda a

¹² PRESNO LINERA, M.A. & MEUWESE, A., *op cit.*, p. 142

nuestra mente sea la de un juez siendo sustituido por un robot capaz de analizar un caso y dictar sentencia de manera automática. Esta idea se ha convertido en un eje recurrente dentro del debate académico sobre inteligencia artificial y justicia teniendo como resultado que numerosos trabajos y artículos hayan abordado esta cuestión, ya sea para explorar sus posibilidades técnicas y teóricas, o para advertir de sus riesgos y limitaciones. En cualquier caso debemos comprender que esta representación, aunque llamativa, resulta simplificadora, reduccionista y no consigue reflejar la complejidad real del fenómeno¹³.

El mundo del derecho es un campo inmenso y heterogéneo, en el que encontramos desde sencillos procedimientos administrativos de escasa complejidad, como solicitar una licencia de pesca, hasta decisiones de enorme trascendencia que pueden afectar directamente a derechos fundamentales, como el procedimiento penal cuyo fallo puede determinar que una persona cumpla una pena de prisión o permanezca en libertad¹⁴. Lo que debemos entender es que cuando hablamos de administración de justicia, nos referimos a un ámbito amplio y complejo que no se reduce únicamente a la toma de decisiones judiciales o a la imposición de sentencias, sino que abarca un conjunto diverso de fases procesales en las que la inteligencia artificial puede desempeñar un papel relevante.

La forma en que estas herramientas de inteligencia artificial se integran en la administración de justicia puede variar enormemente en función de la fase procesal en la que nos encontremos. En ciertos supuestos, los sistemas de IA sustituyen directamente al ser humano en la ejecución de tareas específicas, automatizando funciones que antes requerían intervención manual. Sin embargo, lo más habitual es que estos sistemas actúen como instrumentos de apoyo al juez, al personal judicial o a los abogados, complementando su labor y facilitando la gestión de procesos. En cualquiera de los casos, lo que sí parece

¹³ SOLAR CAYÓN, J.I., “¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la Administración de Justicia”, en: Solar Cayón, José Ignacio y Sánchez Martínez, M^a Olga (dir.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica*, La Ley (Wolters Kluwer), Madrid, 2022, pp. 245-280 (pp. 258-259).

¹⁴ PASQUALE, F., “A rule of persons, not machines: the limits of legal automation”, *The George Washington Law Review*, vol. 87, n^o 1, 2019, pp. 1-55 (p. 6).

claro es que su introducción implica una modificación en los métodos de trabajo tradicionales, generando transformaciones que en ocasiones pueden llegar a ser profundas, tanto en la organización de los tribunales como en el rol que desempeñan los distintos actores del proceso¹⁵.

Con el fin de comprender mejor el alcance de estas aplicaciones, autores como SOLAR CAYÓN¹⁶ proponen clasificar los sistemas de IA en función de la tarea que cumplen dentro del proceso judicial. Desde esta óptica, pueden identificarse tres grandes categorías. En primer lugar, aquellas herramientas de carácter instrumental o auxiliar, que se limitan a tareas de apoyo administrativo o logístico y que no inciden directamente en la decisión judicial. En segundo lugar, encontramos sistemas vinculados a la fase procesal, orientados a facilitar determinados trámites o a proporcionar información relevante que puede influir en la formación del juicio. Finalmente, en una tercera categoría se sitúan aquellas aplicaciones con incidencia directa en la resolución de la disputa, es decir, sistemas capaces de intervenir en la propia decisión judicial, por ejemplo mediante propuestas automatizadas de resolución.

En la tipología propuesta, los sistemas algorítmicos de evaluación del riesgo de reincidencia, sobre los que hablaremos más adelante, se encuadran en el segundo grupo: no deciden la sentencia, sino que apoyan trámites del proceso como la valoración del riesgo en medidas cautelares o libertad provisional.

3.2. Principales ventajas y riesgos

La incorporación de la inteligencia artificial en la administración de justicia abre un campo de reflexión en el que se entrelazan posibilidades de mejora y desafíos significativos. El análisis de sus implicaciones requiere considerar no solo el potencial transformador de estas herramientas, sino también las limitaciones y riesgos que su uso conlleva en un ámbito caracterizado por la necesidad de garantizar imparcialidad, transparencia y respeto a los derechos fundamentales. En este contexto, resulta pertinente examinar de manera equilibrada las principales ventajas y riesgos asociados a la aplicación de la inteligencia artificial

¹⁵ SOLAR CAYÓN, J.I., "Inteligencia artificial y justicia digital", en LLANO ALONSO, F.H. (dir.), *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022, pp. 381-427 (p. 386)

¹⁶ *Ibidem*, pp 386-416

en el sistema judicial, con el objetivo de ofrecer una visión crítica y fundamentada sobre su alcance.

Las funcionalidades que ofrecen las herramientas de inteligencia artificial aplicadas al ámbito jurídico son diversas y abren nuevas perspectivas metodológicas para el ejercicio profesional del derecho. Su incorporación en la práctica de la abogacía, para la cual fueron inicialmente concebidas, ya se ha consolidado en numerosas firmas de gran envergadura que emplean sistemas predictivos disponibles en el mercado o que incluso han desarrollado modelos propios en colaboración con empresas especializadas. Estos instrumentos permiten asistir al jurista en la valoración de aspectos estratégicos fundamentales, como la probabilidad de éxito de una demanda, la conveniencia de litigar o negociar un acuerdo, y la selección de la estrategia procesal más idónea en función de factores como la composición del tribunal¹⁷.

No obstante, el empleo de estas herramientas ha suscitado un intenso debate en torno a sus implicaciones. Mientras algunos sectores sostienen que su uso favorece la transparencia al poner de manifiesto posibles sesgos o tendencias en la actuación judicial, otros advierten de los riesgos que pueden representar para la independencia judicial, como lo demostró la experiencia francesa, donde se limitó por vía legal la utilización de datos personales de jueces con fines de análisis predictivo. Este tipo de medidas busca preservar la autonomía de los magistrados y evitar la mercantilización de estrategias procesales, si bien resulta difícil impedir su aplicación interna en los despachos profesionales¹⁸.

Paralelamente, la discusión sobre el potencial de los sistemas predictivos se extiende también al ámbito institucional, donde se señala su utilidad en el marco de los tribunales en línea instaurados en distintas jurisdicciones, como el *Civil Resolution Tribunal* de Columbia Británica o el *Online Solutions Court* en Inglaterra y Gales. Estas plataformas, concebidas para facilitar el acceso a la justicia y agilizar la resolución de conflictos, integran distintos niveles de actuación que van desde el asesoramiento inicial y la orientación interactiva del

¹⁷ SOLAR CAYÓN, J.I., "La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista", *Revus. Journal for Constitutional Theory and Philosophy of Law*, 41, 2020, pp. 1-27 (p. 13)

¹⁸ SOLAR CAYÓN, J.I., "Inteligencia artificial y justicia digital", cit., p. 383.

usuario, hasta la mediación automatizada y, en última instancia, la decisión judicial remota. En este contexto, los sistemas de análisis predictivo adquieren un papel especialmente relevante en las fases iniciales, al proporcionar a los ciudadanos una herramienta que les permita evaluar con mayor precisión las posibilidades de éxito de sus pretensiones en caso de judicialización, contribuyendo así a una justicia más accesible y orientada a las necesidades de los usuarios¹⁹.

La aplicación de sistemas de inteligencia artificial en el ámbito judicial plantea un debate particularmente complejo cuando se considera su posible utilización para la toma de decisiones jurisdiccionales. Resulta necesario distinguir entre el empleo de estas herramientas por parte de abogados o usuarios, orientados a estimar la viabilidad de una estrategia procesal o de una pretensión concreta, y su integración en el proceso decisorio de los jueces. Si bien los sistemas predictivos pueden ofrecer estimaciones valiosas en función de precedentes consolidados, la labor judicial no se limita a predecir, sino a decidir y, sobre todo, a justificar normativamente una resolución. A ello se añade el riesgo de los denominados modelos de “caja negra”, propios del aprendizaje profundo, cuya opacidad resulta incompatible con la exigencia de motivación y transparencia inherente a una función pública de esta naturaleza ²⁰.

Pese a estas limitaciones, algunos plantean escenarios pragmáticos en los que, ante la saturación de determinados sistemas judiciales, el uso de predicciones automatizadas podría concebirse como un mal menor frente a la parálisis institucional, especialmente en casos rutinarios o de escaso valor. No obstante, experiencias recientes, como el sistema “*Prometea*” en Argentina o “*PretorIA*” en Colombia, demuestran que el uso más prometedor de la inteligencia artificial en este terreno no reside en sustituir al juez, sino en asistirlo. CORVALÁN, J.G. deja evidencia de esto explicando cómo “*Prometea*” fue entrenada para analizar y detectar casos prioritarios en materia de salud y tras unos meses de preparación “pudo seleccionar 32 casos prioritarios en 2 minutos, lo que a una persona le

¹⁹ SOLAR CAYÓN, J.I., “La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista”, cit., pp. 13-14.

²⁰ CASTILLEJO MANZANARES, R., “Digitalización y/o Inteligencia Artificial”, en CALAZA LÓPEZ, S. y LLORENTE SÁNCHEZ-ARJONA, M. (dir.), *Inteligencia Artificial Legal y Administración de Justicia*, Aranzadi, Cizur Menor (Navarra), 2022 pp. 55-90 (p. 79-81)

llevaría 96 días laborales”²¹. Estos modelos, diseñados con metodologías transparentes y bajo un marco institucional que preserva la competencia humana, han mostrado un alto grado de fiabilidad en la agilización de procedimientos y en la clasificación de casos, sin menoscabar la legitimidad de la decisión judicial. En consecuencia, la cuestión no radica en la viabilidad inmediata de un “juez-robot”, sino en el desarrollo de herramientas capaces de complementar la labor judicial de manera transparente, trazable y sometida a control humano. Ello exige un debate profundo y sostenido que permita definir con claridad las posibilidades y límites de estas tecnologías, de modo que se aprovechen sus potencialidades sin comprometer los principios esenciales de independencia, motivación y legitimidad que sustentan la función jurisdiccional²².

4. SISTEMAS ALGORÍTMICOS DE EVALUACIÓN DE RIESGOS DE REINCIDENCIA CRIMINAL

Dentro del uso de la inteligencia artificial en la administración de justicia, los sistemas algorítmicos de evaluación de riesgos de reincidencia criminal se han consolidado como un ejemplo paradigmático de la incorporación de estas tecnologías en el ámbito judicial y penitenciario. Su creciente presencia ha suscitado un intenso debate académico y práctico, al situarse en la intersección entre la búsqueda de eficiencia en la gestión judicial y penitenciaria y la necesidad de garantizar los principios de imparcialidad, transparencia y respeto a los derechos fundamentales.

²¹ CORVALÁN, J. G., “El impacto de la Inteligencia Artificial en el trabajo”, *Revista de Direito Econômico e Socioambiental*, Vol. 10 Núm. 1, 2019, pp. 35-51 (pp. 44-45) DOI: 10.7213/rev.dir.econ.soc.v10i1.25870

²² SOLAR CAYÓN, J.I., “La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista”, cit., p. 15.

Este tipo de sistemas están fundamentados en técnicas de *machine learning* y en el análisis masivo de datos históricos, a través de los cuales buscan estimar la probabilidad de que un individuo vuelva a cometer un delito en el futuro. El origen de estos tiene lugar en el ámbito penitenciario, como una evolución de los modelos estadísticos tradicionales que, desde hace décadas, venían empleándose para valorar la probabilidad de reincidencia de los reclusos. Inicialmente, la finalidad de esta herramienta era la de orientar a las comisiones competentes de los centros penitenciarios para asistirles en su labor de diseñar programas de tratamiento y rehabilitación personalizados para las necesidades de cada reo. Además podían servir para ofrecer criterios adicionales que ayudasen a decidir sobre la concesión de la libertad condicional, así como el grado de supervisión correspondiente²³.

No obstante, a lo largo de los últimos años, estos sistemas han evolucionado hasta trascender su marco original y paulatinamente incorporarse al proceso judicial, especialmente en la fase del *pre-trial*, es decir, la fase previa al juicio. En este contexto, dichos sistemas se aplican para apoyar las decisiones relativas a medidas cautelares, como la prisión preventiva, la fijación de fianzas o la determinación de las condiciones de la libertad provisional. Hemos de entender que en esta fase del proceso, estos sistemas no buscan sustituir al juez, sino más bien, orientarle y proporcionarle información complementaria que ayude a tomar las medidas cautelares que sean más acordes al caso concreto.²⁴

La creciente utilización de este tipo de sistemas en jurisdicciones como la estadounidense pone de relieve, no solo el potencial de la inteligencia artificial para optimizar la gestión judicial y penitenciaria, sino también la necesidad de llevar a cabo un examen crítico acerca de su fiabilidad, transparencia y compatibilidad con los principios fundamentales que rigen el derecho penal y procesal.

²³ SOLAR CAYÓN, J.I., "Inteligencia artificial y justicia digital", cit., p. 403.

²⁴ SOLAR CAYÓN, J.I., "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos", en SOLAR CAYÓN, J.I (ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Universidad de Alcalá - Defensor del Pueblo, Alcalá de Henares, 2020, pp. 125-172 (p. 130)

4.1. Funcionamiento básico

Como hemos mencionado, estos sistemas funcionan mediante machine learning, sin embargo, aún no nos hemos detenido a explicar en qué consiste este campo. El *machine learning* es una rama de la Inteligencia Artificial que busca dotar a los sistemas informáticos de la capacidad de aprender de manera autónoma a partir de datos de entrenamiento, emulando el proceso de aprendizaje humano. A través de algoritmos diseñados para generar modelos matemáticos, es decir, nuevos algoritmos, esta tecnología permite analizar bases de datos, identificar patrones, reconocer similitudes y adaptarse de forma progresiva al procesar nueva información, lo que posibilita realizar predicciones, tomar decisiones y resolver problemas complejos²⁵.

En este sentido, el *machine learning* constituye una herramienta fundamental para el análisis de grandes volúmenes de datos, ya que no solo facilita la detección de comportamientos ocultos, sino que también contribuye a la automatización de tareas y a la optimización de procesos en diferentes contextos.

En relación con la evaluación de riesgos de reincidencia criminal, estos sistemas recurren a modelos estadísticos que identifican patrones y correlaciones entre determinados factores personales, sociales y judiciales (como edad, nivel educativo, entorno familiar, situación laboral, consumo de estupefacientes, lugar de residencia, contexto socioeconómico o antecedentes penales), asignando a cada uno de ellos un peso específico en la configuración del perfil de riesgo. De este modo, los algoritmos generan predicciones que clasifican a los sujetos en distintos niveles de peligrosidad, habitualmente definidos como riesgo bajo, medio o alto²⁶.

²⁵ CASTILLEJO MANZANARES, R., "Digitalización y/o Inteligencia Artificial", en CALAZA LÓPEZ, S. y LLORENTE SÁNCHEZ-ARJONA, M. (dir.), *Inteligencia Artificial Legal y Administración de Justicia*, Aranzadi, Cizur Menor (Navarra), 2022 pp. 55-90 (p. 71)

²⁶ SOLAR CAYÓN, J.I., "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos", cit., p. 139.

4.2. Principales modelos existentes: COMPAS en EE.UU.

En el ámbito del proceso penal contemporáneo se emplean diversos algoritmos predictivos de riesgo, cada uno con finalidades y criterios específicos. ROA AVELLA, M., SANABRIA-MOYANO, J. y DINAS-HURTADO, K.²⁷, comentan en qué consisten los más importantes, los cuales mencionaremos a continuación.

En primer lugar, el *Public Safety Assessment* (PSA) se utiliza para establecer si una persona procesada puede o no seguir en libertad hasta el comienzo del juicio, considerando factores asociados al riesgo de reincidencia o de incomparecencia.

En segundo lugar, el *Level of Service Inventory-Revised* (LSI-R) interviene en la concesión de permisos de salida y de libertad condicional, tomando en cuenta aspectos tales como los antecedentes penales, el entorno familiar, el nivel educativo, la situación laboral, los hábitos de ocio, así como posibles problemas relacionados con el alcoholismo, la drogadicción u otras características personales.

En tercer lugar, el *Harm Assessment Risk Tool* (HART) tiene como fin calcular la probabilidad de que un sujeto cometa un delito en el plazo de dos años. Para ello emplea variables como la edad, el sexo y el historial delictivo; en caso de que un individuo presente una probabilidad baja de cometer un delito, podría determinarse su envío a un programa de rehabilitación.

En cuarto lugar, el *Crime Anticipation System* (CAS) busca prever la comisión de delitos de alto impacto, tales como el hurto, en zonas concretas. Para ello, combina datos históricos sobre crímenes ocurridos en el lugar con información socioeconómica y demográfica (edad, género, etnia, nivel de ingresos, residencia) e incluso con la proximidad geográfica de posibles sospechosos, siguiendo principios de la criminología ambiental.

²⁷ ROA AVELLA, M., SANABRIA-MOYANO, J., DINAS-HURTADO, K., "Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos", *Revista Brasileira de Direito Processual*, v. 8, n. 1, pp. 275-310 (pp. 280-281) DOI: <https://doi.org/10.22197/rbdpp.v8i1.615>

Finalmente, el *National Data Analytics Solution* (NDAS) se centra en la predicción de delitos violentos, en particular aquellos cometidos con armas blancas o de fuego, identificando tanto a posibles perpetradores como a potenciales víctimas. Este sistema emplea registros de detenciones y delitos previos, y a partir de la identificación de individuos en situación de riesgo, promueve la oferta de servicios sociales y de salud orientados a prevenir la materialización de estas conductas delictivas.

Como podemos comprobar, son varios los sistemas algorítmicos utilizados en este ámbito, sin embargo, en este trabajo, nos centraremos principalmente en el sistema COMPAS puesto que se ha convertido en uno de los más utilizados y debatidos en Estados Unidos. Su amplia implementación y el impacto directo que tiene en decisiones legales, como la libertad condicional o la fijación de penas, lo convierten en un referente clave para analizar los alcances, limitaciones y controversias éticas vinculadas al uso de algoritmos en la predicción del riesgo criminal.

El sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) fue diseñado por la empresa privada Northpointe como una herramienta de evaluación de riesgos utilizada en el ámbito penal. Según sus creadores, se trata de un instrumento de “cuarta generación”, orientado a realizar un seguimiento continuo de los individuos mediante el análisis de factores de riesgo dinámicos, con el objetivo de apoyar las políticas de intervención correccional y reducir la probabilidad de reincidencia delictiva²⁸.

El algoritmo incorpora dos modelos principales de predicción: por un lado, el riesgo general de reincidencia; por otro, la probabilidad específica de reincidencia en conductas violentas. Para ello, integra tanto factores dinámicos, vinculados a características personales o contextuales susceptibles de cambio (por ejemplo, entorno social o consumo de sustancias), como factores estáticos, referidos a circunstancias invariables, como los antecedentes penales. En la práctica, COMPAS se emplea en el proceso penal para ofrecer una estimación del grado de riesgo que presenta un acusado. La valoración se construye a partir

²⁸ NORTHPOINTE, *Practitioner's Guide to COMPAS*, Core, 2015

de un cuestionario extenso (compuesto por más de un centenar de ítems agrupados en distintas secciones) y de los registros sobre el historial criminal del sujeto, complementados habitualmente con entrevistas personales. Toda esta información se somete a un proceso de ponderación algorítmica que genera un resultado cuantitativo: un “score” de riesgo en una escala de 1 a 10, donde los valores más bajos indican menor probabilidad de reincidencia y los más altos reflejan un mayor nivel de peligrosidad futura²⁹.

5. PRINCIPALES PROBLEMAS QUE PLANTEA SU USO

A la hora de analizar el sistema COMPAS, resulta ineludible no mencionar el caso *State of Wisconsin v. Loomis*³⁰, el cual es considerado como un precedente significativo, al ser una de las primeras ocasiones en que un tribunal estadounidense se pronunció sobre la pertinencia de utilizar herramientas algorítmicas en el proceso penal, resolviendo que el uso de estos sistemas eran constitucionales, siempre y cuando se cumpliesen ciertos requisitos: principalmente, el carácter orientativo del resultado que indique COMPAS; además de la prohibición expresa de que el fallo judicial se base únicamente en dichos resultados.

En el mencionado asunto, el acusado, Eric Loomis, enfrentaba inicialmente cinco cargos vinculados a un tiroteo, entre ellos posesión ilícita de armas de fuego, posesión de un rifle de cañón corto, intento de eludir a un agente de tráfico,

²⁹ ROAAVELLA, M., SANABRIA-MOYANO, J., DINAS-HURTADO, K., *op cit.*, pp. 282-283.

³⁰ *State v. Loomis*, 881 N.W. 2d 749 (Wis. 2016).

conducción de un vehículo sin consentimiento del propietario y puesta en peligro de la seguridad pública. Aunque negó haber participado en el tiroteo, reconoció haber conducido el automóvil implicado e intentado huir de la policía. Finalmente, alcanzó un acuerdo con la Fiscalía por el cual admitió únicamente los delitos de conducción no autorizada y puesta en peligro de la seguridad pública. Durante la fase de determinación de la pena, la fiscalía presentó un informe elaborado con el sistema COMPAS, en el que se advertía que el acusado presentaba un elevado riesgo de reincidencia. El tribunal de primera instancia, tomando en consideración dicho informe, junto con otros elementos, condenó a Loomis a seis años de prisión y a cinco años adicionales de libertad vigilada³¹.

La defensa recurrió la sentencia ante la Corte Suprema de Wisconsin, alegando la vulneración del derecho al debido proceso. Los aspectos más problemáticos de estas aplicaciones fueron reconocidos por el Tribunal Supremo de Wisconsin y por ello, procederemos a comentarlos a continuación.

5.1. Dificultades de transparencia

La opacidad de los sistemas de inteligencia artificial en el ámbito judicial puede tener distintos orígenes. En algunos casos, responde a factores jurídicos vinculados a la protección del secreto empresarial, como ocurre con el algoritmo COMPAS, desarrollado por una empresa privada. Este tipo de situaciones podría afrontarse mediante diversas medidas, tales como establecer procedimientos que permitan a la defensa o a peritos especializados acceder al algoritmo bajo condiciones de confidencialidad, exigir la renuncia parcial al secreto como requisito para comercializar estas herramientas en la Administración de Justicia, o recurrir a soluciones basadas en software de código abierto. No obstante, la falta de transparencia puede derivar también de condicionantes técnicos, especialmente en los sistemas basados en redes neuronales profundas. A diferencia de los modelos algorítmicos tradicionales, el *deep learning* genera sus propias estructuras predictivas a partir de enormes volúmenes de datos, ajustando constantemente sus parámetros en función de la experiencia acumulada. Ello convierte a estos sistemas en auténticas “cajas negras”, donde

³¹ ROAAVELLA, M., SANABRIA-MOYANO, J., DINAS-HURTADO, K., *op cit.*, pp. 287-288.

ni siquiera el acceso al código fuente permite identificar con claridad qué peso ha tenido cada variable en el resultado final. Este grado de opacidad plantea serias dudas sobre su idoneidad en el proceso penal, pues dificulta la posibilidad de someter a contradicción los resultados y amenaza garantías esenciales como el derecho de defensa³².

5.2. Imparcialidad y sesgos algorítmicos

Uno de los desafíos más controvertidos en el uso de sistemas algorítmicos en el ámbito penal es el de los sesgos. El caso de COMPAS se ha convertido en un ejemplo paradigmático a raíz de la investigación publicada por la ONG *ProPublica*, que puso de relieve la existencia de un posible sesgo racial en sus predicciones. Según este estudio, el sistema tendía a clasificar de manera desproporcionada a los acusados afroamericanos como individuos de alto riesgo de reincidencia, mientras que en el caso de los acusados blancos se producía el fenómeno inverso: eran más frecuentemente catalogados como de bajo riesgo. Frente a estas críticas, la empresa desarrolladora de COMPAS respondió con un informe en el que cuestionaba la metodología utilizada por *ProPublica*, señalando la ausencia de consideraciones estadísticas relevantes, como las diferencias en las tasas base de reincidencia entre los distintos grupos raciales, así como la omisión de métricas reconocidas de equidad algorítmica, como la paridad predictiva o la calibración, que en su opinión el sistema sí cumplía. Más allá de este debate técnico, lo cierto es que la controversia puso de relieve un problema estructural: los algoritmos aprenden a partir de datos históricos y, en consecuencia, tienden a reproducir los sesgos ya presentes en el sistema judicial y policial. De hecho, incluso aceptando que las tasas de predicción fueran equivalentes entre los distintos grupos, ello no elimina el riesgo de que el algoritmo consolide desigualdades preexistentes. Algo similar ocurre con el tratamiento diferenciado que COMPAS aplica según el género, dado que las mujeres presentan históricamente menores tasas de reincidencia que los hombres. En este caso, el propio tribunal avaló esta distinción, entendiendo que un enfoque aparentemente neutro conduciría a resultados injustos. Este ejemplo

³² SOLAR CAYÓN, J.I., “Inteligencia artificial y justicia digital”, cit., pp. 404-405.

ilustra la complejidad de conciliar exigencias técnicas de precisión con los principios jurídicos de igualdad y no discriminación³³.

5.3. Uso de “perfiles” para decisiones individualizadas: impacto en la presunción de inocencia

El uso de algoritmos de evaluación de riesgos en el ámbito penal plantea también la problemática del perfilado, ya que estos sistemas no valoran exclusivamente las circunstancias concretas del individuo evaluado, sino que se apoyan en datos estadísticos de grupos poblacionales que comparten determinados atributos. Este fenómeno se inscribe dentro de lo que se conoce como *justicia actuarial*, cuyo rasgo esencial, como señala DOMINIQUE ROBERT³⁴, consiste en transformar fenómenos individuales y sociales en factores de riesgo. En consecuencia, el foco deja de estar en la persona considerada en su singularidad biográfica, para situarse en la construcción de un perfil probabilístico. La identidad del sujeto se descompone en variables asociadas a diferentes categorías y niveles de riesgo, configurando así un modelo algorítmico que no describe de manera fiel a ningún individuo en particular, sino que resulta de la combinación de múltiples atributos derivados del análisis de grandes bases de datos. De este modo, se asigna a un acusado un determinado nivel de peligrosidad únicamente por compartir ciertos rasgos con un colectivo previamente identificado como de riesgo, lo que suscita serias dudas en relación con principios jurídicos como la individualización de la responsabilidad penal³⁵.

En definitiva, los riesgos asociados a los sistemas algorítmicos de evaluación de reincidencia no son meras conjeturas teóricas, sino cuestiones reales y de gran trascendencia, como reconoció expresamente el Tribunal Supremo de Wisconsin en el caso *Loomis*. Estas advertencias ponen de relieve la necesidad de no trivializar ni pasar por alto tales problemas, pues afectan directamente a derechos fundamentales y a la legitimidad del proceso penal. De ahí que resulte imprescindible mantener un análisis constante y una evaluación continua de

³³ SOLAR CAYÓN, J.I., “Inteligencia artificial y justicia digital”, cit., pp. 405-406.

³⁴ ROBERT, D., “Actuarial Justice”, en BOSWORTH, M. (ed.), *Encyclopedia of Prisons & Correctional Facilities*, Sage, London, 2005, pp. 11-13.

³⁵ SOLAR CAYÓN, J.I., “Inteligencia artificial y justicia digital”, cit., p. 405.

estas herramientas, garantizando que su utilización se someta siempre a un escrutinio crítico y riguroso.

6. COMPATIBILIDAD CON EL MARCO JURÍDICO EUROPEO

La regulación de la inteligencia artificial en la Unión Europea se estructura a partir de un enfoque basado en el riesgo, lo que implica que la intensidad de las exigencias normativas se adecúe al grado de amenaza potencial que cada sistema pueda generar para la salud, la seguridad o los derechos fundamentales de las personas. Este planteamiento se vincula estrechamente con el principio de precaución, que constituye uno de los ejes rectores de la actuación comunitaria y que ya había sido invocado en resoluciones del Parlamento Europeo en relación con la robótica y las tecnologías emergentes. De acuerdo con esta lógica, algunos sistemas de IA se consideran inadmisibles por el nivel de riesgo que comportan, lo que conduce directamente a su prohibición, mientras que otros se clasifican como de alto riesgo, debiendo someterse a rigurosos mecanismos de gestión y control a lo largo de todo su ciclo de vida. Dichos mecanismos abarcan la identificación y análisis de riesgos previsibles, la evaluación de posibles escenarios de uso indebido, la incorporación de datos derivados de la vigilancia posterior a su comercialización y la adopción de medidas correctivas específicas. Además, la normativa tiene en cuenta factores contextuales y las competencias técnicas de los responsables de su implementación, con el fin de garantizar que la aplicación de la IA se produzca

en condiciones de seguridad, transparencia y respeto a los valores fundamentales de la Unión³⁶.

En el marco del Reglamento (UE) 2024/1689, los sistemas de evaluación algorítmica del riesgo de reincidencia reciben un tratamiento diferenciado en función de sus características. Por un lado, el Anexo III, apartado 6(d) contempla la posibilidad de emplear este tipo de sistemas “en apoyo a las autoridades garantes del cumplimiento del derecho” siempre que no se basen únicamente en la “elaboración de perfiles de personas físicas mencionada en el artículo 3, punto 4, de la Directiva (UE) 2016/680 o para evaluar rasgos y características de la personalidad o comportamientos delictivos pasados de personas físicas o colectivos” y que su uso se encuentre amparado por el Derecho de la Unión o por el ordenamiento nacional aplicable. En este caso, los sistemas son considerados de alto riesgo, lo que conlleva la obligación de ajustarse a las exigencias previstas en el capítulo III del Reglamento, entre ellas la gestión de riesgos, la gobernanza de los datos, la supervisión humana y, de manera destacada, el cumplimiento de niveles adecuados de transparencia (tal y como se indica en el art. 13 del reglamento), lo que plantea serias dificultades en el caso de algoritmos basados en aprendizaje profundo, dada su naturaleza opaca. Además, los proveedores de estas tecnologías deberán ofrecer información detallada sobre los parámetros estadísticos que avalan su fiabilidad y capacidad predictiva, el tamaño y características de la población utilizada para su entrenamiento, las variables seleccionadas como factores de riesgo, los métodos de ponderación aplicados y la duración temporal de la validez de las predicciones. Igualmente, deberán hacerse públicos los estudios de validación que sustenten la eficacia del sistema, de modo que se garantice un control adecuado sobre su funcionamiento y sus posibles limitaciones³⁷.

Por otro lado, en caso de que estos sistemas se basen únicamente en la elaboración de perfiles mencionados, pasaría a ser de aplicación el artículo 5, apartado d), el cual establece su prohibición. Dicho artículo recalca que la predicción del riesgo no puede fundamentarse “únicamente en la elaboración del

³⁶ PRESNO LINERA, M.A. & MEUWESE, A., *op cit.*, pp. 145-147

³⁷ SOLAR CAYÓN, J.I., “Inteligencia artificial y justicia digital”, cit., p. 407.

perfil de una persona física o en la evaluación de los rasgos y características de su personalidad”. En estos supuestos, el legislador europeo considera que la afectación a los derechos fundamentales es de tal magnitud que impide cualquier margen de utilización legítima salvo cuando dichos sistemas sean “utilizados para apoyar la valoración humana de la implicación de una persona en una actividad delictiva”.

Esta distinción normativa pone de relieve que el legislador europeo no niega en bloque la utilidad de estas herramientas, pero sí impone un marco especialmente riguroso de garantías y controles con el fin de reducir al mínimo sus potenciales efectos negativos.

7. CONCLUSIONES

El análisis realizado a lo largo de este Trabajo de Fin de Grado permite extraer una serie de reflexiones en torno al uso de sistemas algorítmicos de evaluación de riesgos de reincidencia criminal y, en general, sobre la incorporación de la inteligencia artificial (IA) en el ámbito de la Administración de Justicia.

En primer lugar, resulta innegable que la irrupción de la inteligencia artificial ha transformado de manera significativa el panorama jurídico contemporáneo. Los sistemas de justicia, tradicionalmente caracterizados por procedimientos lentos, rígidos y fuertemente dependientes de la intervención humana, se enfrentan a la oportunidad de modernizarse a través de herramientas capaces de procesar grandes volúmenes de datos y generar información relevante para la toma de decisiones. En este sentido, los sistemas algorítmicos de evaluación del riesgo de reincidencia representan un ejemplo paradigmático: pretenden aportar objetividad, eficiencia y racionalización en un campo especialmente sensible como es el derecho penal.

Ahora bien, la investigación muestra que el uso de estos sistemas no está exento de problemas. El caso de COMPAS, ampliamente empleado en Estados Unidos, pone de relieve una serie de riesgos que no pueden pasarse por alto. Entre ellos destacan la falta de transparencia de los algoritmos, la posibilidad de sesgos

discriminatorios (particularmente de carácter racial o de género), así como las tensiones que generan respecto de principios nucleares del derecho penal, como la presunción de inocencia o la individualización de la responsabilidad criminal. El precedente judicial de *State v. Loomis* evidenció la necesidad de fijar límites claros: las valoraciones algorítmicas no pueden ser el único fundamento de una sentencia, sino que deben tener un carácter meramente orientativo.

A pesar de estas advertencias, sería un error desestimar de plano la utilidad de estas herramientas. La cuestión, como hemos tratado de demostrar, no radica tanto en la viabilidad técnica de los sistemas, sino en la manera en que se integran dentro del proceso judicial y en el marco normativo que se establezca para su funcionamiento. El análisis del marco europeo es especialmente ilustrativo: el Reglamento (UE) 2024/1689 adopta un enfoque basado en el riesgo, clasificando a los sistemas de evaluación de reincidencia como de “alto riesgo” y sometiéndolos a exigencias estrictas en materia de transparencia, fiabilidad, supervisión humana y protección de derechos fundamentales. Esta aproximación no prohíbe el uso de estas herramientas, pero condiciona su aplicación a un riguroso control, lo cual representa una solución equilibrada entre innovación tecnológica y garantía jurídica.

La compatibilidad de la IA con el marco jurídico europeo dependerá, en última instancia, de la capacidad de los Estados miembros y de las instituciones comunitarias para desarrollar mecanismos de control efectivos. Ello requiere no solo disposiciones normativas claras, sino también medios técnicos y humanos suficientes para auditar los algoritmos, garantizar la trazabilidad de sus decisiones y evaluar de forma independiente su fiabilidad. El reto es mayúsculo, puesto que exige una combinación de competencias jurídicas, informáticas, estadísticas y éticas que hasta ahora no habían convergido en el ámbito judicial.

Otra conclusión relevante es que los riesgos identificados en los sistemas algorítmicos no son, en esencia, completamente novedosos. Muchos de los problemas que se atribuyen a la IA (falta de transparencia, sesgos, desigualdades estructurales) ya existían previamente en la actuación de jueces, fiscales y policías. La diferencia es que, cuando dichos sesgos se trasladan a un

sistema automatizado, existe el peligro de consolidarlos de manera opaca y a gran escala. Sin embargo, precisamente por tratarse de sistemas programados, existe la posibilidad de someterlos a controles objetivos más estrictos que los aplicados a la discrecionalidad humana. Esta circunstancia abre la puerta a una concepción optimista: en lugar de eliminar la imparcialidad judicial, los algoritmos, bien diseñados y auditados, podrían contribuir a visibilizar y reducir los sesgos preexistentes.

Asimismo, debe subrayarse que el objetivo último de estos sistemas no es, ni debe ser, sustituir al juez humano, sino asistirle en la toma de decisiones. Como muestran las experiencias de proyectos en Latinoamérica (*Prometea*, *PretorIA*), la IA puede convertirse en un instrumento valioso para agilizar trámites, detectar casos prioritarios y optimizar recursos, sin desplazar el papel de la persona como garante de los principios de legitimidad, motivación y justicia material. En este sentido, los sistemas de evaluación de riesgo no deben entenderse como oráculos infalibles, sino como instrumentos de apoyo que enriquecen la deliberación judicial siempre que su funcionamiento se someta a estándares claros de transparencia y control.

Un aspecto adicional que merece destacarse es el carácter ineludiblemente político de estas tecnologías. Los algoritmos no son neutrales: están diseñados por empresas o instituciones que seleccionan variables, fijan ponderaciones y deciden qué datos son relevantes. En consecuencia, la discusión sobre su uso en la justicia penal no puede reducirse a una cuestión meramente técnica, sino que debe abordarse desde una perspectiva democrática, en la que participen no solo juristas y programadores, sino también ciudadanos, organizaciones sociales y colectivos potencialmente afectados. Solo así se garantiza que el diseño e implementación de estas herramientas respondan al interés general y no exclusivamente a lógicas de eficiencia administrativa o rentabilidad empresarial.

Finalmente, y como reflexión última, me gustaría romper una lanza a favor del potencial de la inteligencia artificial en la Administración de Justicia. Como afirma Ben Green, “with appropriate technical assurances of fairness, risk assessments that inform bail and sentencing decisions can be tailored into neutral tools to

improve the criminal justice system”³⁸. Este planteamiento encierra una idea fundamental: la tecnología, por sí sola, no es buena ni mala; depende de cómo se diseñe, se regule y se aplique. Si logramos garantizar la equidad técnica, la supervisión humana y el respeto a los derechos fundamentales, los sistemas de evaluación de riesgos pueden convertirse en instrumentos neutrales que, lejos de agravar las deficiencias del sistema penal, lo mejoren.

La clave reside en concebir la inteligencia artificial no como un sustituto del juez, sino como un aliado. Un aliado que aporta datos, estadísticas y predicciones que el ser humano difícilmente podría procesar por sí solo, pero que sigue necesitando de la mirada crítica y del juicio normativo del profesional del derecho. Solo a través de esta sinergia podremos avanzar hacia un modelo de justicia verdaderamente digital, pero también profundamente humano.

En conclusión, los sistemas algorítmicos de evaluación de riesgos de reincidencia criminal representan un desafío complejo, pero también una oportunidad única. Si se adoptan con cautela, transparencia y bajo marcos normativos sólidos, pueden contribuir a una justicia más eficiente, más coherente y, en última instancia, más justa. Renunciar a ellos por miedo a sus riesgos sería desaprovechar el potencial transformador de la inteligencia artificial. Lo verdaderamente relevante es aprender a utilizarlos con responsabilidad, de modo que se conviertan en herramientas al servicio de los principios que inspiran nuestro Estado de Derecho.

³⁸ GREEN, B., “Fair Risk Assessments: A Precarious Approach for Criminal Justice Reform”, 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018), Stockholm, 2018, p. 1.

8. BIBLIOGRAFÍA

CASTILLEJO MANZANARES, R., “Digitalización y/o Inteligencia Artificial”, en CALAZA LÓPEZ, S. y LLORENTE SÁNCHEZ-ARJONA, M. (dir.), *Inteligencia Artificial Legal y Administración de Justicia*, Aranzadi, Cizur Menor (Navarra), 2022 pp. 55-90.

CORVALÁN, J. G., “El impacto de la Inteligencia Artificial en el trabajo”, *Revista de Direito Econômico e Socioambiental*, Vol. 10 Núm. 1, 2019, pp. 35-51 DOI: 10.7213/rev.dir.econ.soc.v10i1.25870

CRIADO, J.I., “Inteligencia Artificial (y Administración Pública)”, *Eunomia. Revista en Cultura de la Legalidad*, 20, 2021, pp. 348-372. DOI: <https://doi.org/10.20318/eunomia.2021.6097>

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, Council of Europe CEPEJ (2018).

GREEN, B., “Fair Risk Assessments: A Precarious Approach for Criminal Justice Reform”, 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018), Stockholm, 2018.

NORTHPOINTE, Practitioner’s Guide to COMPAS, Core, 2015.

OCDE. (2020). *Legal recommendation 0449*, 22 de mayo de 2019. Recuperado de: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

PASQUALE, F., "A rule of persons, not machines: the limits of legal automation", *The George Washington Law Review*, vol. 87, nº 1, 2019, pp. 1-55.

PRESNO LINERA, M.A, & MEUWESE, A., "La regulación de la inteligencia artificial en Europa", *Teoría y Realidad Constitucional*, (54), 2024, pp. 131–161 DOI: <https://doi.org/10.5944/trc.54.2024.43310>

"*Qué es la Inteligencia Artificial*", Plan de Recuperación, Transformación y Resiliencia, 19 de abril de 2023. Disponible en: <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr> (consultado el 12 de julio de 2025).

ROA AVELLA, M., SANABRIA-MOYANO, J., DINAS-HURTADO, K., "Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos", *Revista Brasileira de Direito Processual*, v. 8, n. 1, 2022 pp. 275-310 DOI: <https://doi.org/10.22197/rbdpp.v8i1.615>

ROBERT, D., "Actuarial Justice", en: BOSWORTH, M. (ed.), *Encyclopedia of Prisons & Correctional Facilities*, Sage, London, 2005, pp. 11-13.

RUSSELL, S., & NORVIG, P., *Artificial Intelligence: A Modern Approach*, 4th Global Ed., Pearson Global Editions, 2021.

SOLAR CAYÓN, J.I., "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos", en SOLAR CAYÓN, J.I (ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Universidad de Alcalá - Defensor del Pueblo, Alcalá de Henares, 2020, pp. 125-172.

SOLAR CAYÓN, J.I., "La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista", *Revus. Journal for Constitutional Theory and Philosophy of Law*, 41, 2020, pp. 1-27.

SOLAR CAYÓN, J. I., “Reflexiones sobre la aplicación de la inteligencia artificial en la administración de justicia”. *Teoría Jurídica Contemporânea*, vol. 6, 2021. Disponible en <https://repositorio.unican.es/xmlui/handle/10902/24149>.

SOLAR CAYÓN, J.I., “Inteligencia artificial y justicia digital”, en LLANO ALONSO, F.H. (dir.), *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022, pp. 381-427.

SOLAR CAYÓN, J.I., “¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la Administración de Justicia”, en: Solar Cayón, José Ignacio y Sánchez Martínez, M^a Olga (dir.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica*, La Ley (Wolters Kluwer), Madrid, 2022, pp. 245-280.

State v. Loomis, 881 N.W. 2d 749 (Wis. 2016).

TURING, A. M., “Computing Machinery and Intelligence”, *Mind*, volume LIX, Issue 236, October 1950, pp. 433-460. DOI: <https://doi.org/10.1093/mind/lix.236.433>