FACULTAD DE CIENCIAS HUMANAS Y SOCIALES

Algoritmos y sistemas de inteligencia artificial como herramienta predictiva del delito: oportunidades, riesgos y límites ético-legales.

Autor/a: Isabel Guillén Lanzarote

Director/a: Mario Juy

Madrid

2025/2026

ÍNDICE

RESUMEN

El presente Trabajo de Fin de Grado analiza el uso de algoritmos y sistemas de inteligencia artificial como herramientas predictivas del delito, valorando sus oportunidades, riesgos y límites ético-legales desde la perspectiva criminológica y jurídico-penal. A partir de una revisión bibliográfica cualitativa y del estudio de casos internacionales relevantes, se examina la evolución histórica de los modelos actuariales hasta los sistemas de policía predictiva y evaluación de riesgo.

El trabajo pone de manifiesto que la aplicación de estas tecnologías puede contribuir a la optimización de recursos policiales, la prevención del delito y el apoyo en la toma de decisiones judiciales, especialmente cuando se integran como herramientas complementarias al juicio humano.

ABSTRACT

This Final Degree Project analyzes the use of algorithms and artificial intelligence systems as crime prediction tools, assessing their opportunities, risks and ethical-legal limitations from criminological and law perspective. Through a qualitative literature review and the studies of relevant international cases, the study explores the evolution from actuarial models to contemporary predictive policing and risk assessment systems.

The research that technologies may contribute to the optimization of police resources, crime prevention and support for judicial decision-making, especially when integrated as complementary tools to human judgment.

1 Introducción

1.1 Presentación del tema

El creciente uso de algoritmos y sistemas de inteligencia artificial en la predicción del delito plantea un desafío de notable significación para el derecho penal, la criminología y la sociedad en conjunto.

Desde la perspectiva criminológica , este fenómeno ha generado especial interés por su amplia utilidad como herramienta de apoyo en la prevención del delito con el objetivo evaluación el riesgo de reincidencia (Brantingham, Valasik & Mohler, 2018). No obstante, resulta esencial que estos se consideren complementarios a los análisis criminológicos, y no sustitutos de ellos. Cabe mencionar, que resulta primordial examinar la compatibilidad de estos sistemas con los principios constitucionales que rigen el derecho penal español.

El presente trabajo busca, analizar las oportunidades, riesgos y límites ético-legales derivados del uso de la inteligencia artificial como herramienta predictiva en el ámbito criminológico y penal.

1.2 Justificación del tema

El creciente uso de algoritmos y sistemas de inteligencia artificial en la predicción del delito constituye una cuestión de especial relevancia para la criminología y el derecho penal debido a su impacto en la toma de decisiones preventivas y judiciales. El uso en distintos contextos de herramientas como PredPol y COMPAS ha generado un amplio debate sobre su eficacia y sobre sus posibles riesgos asociados a la reproducción de sesgos y discriminaciones derivadas de datos utilizados. Estos sistemas se han convertido en referentes para el análisis de las implicaciones éticas, sociales y jurídicas en la predicción algorítmica del delito.

Desde la perspectiva jurídico-penal, resulta necesario examinar la compatibilidad de estas tecnologías con los principios constitucionales y los derechos fundamentales, además de las exigencias de transparencia, control humano y rendición de cuentas señaladas por las instituciones europeas. En este contexto, la justificación del presente trabajo reside en la necesidad de promover un análisis crítico e interdisciplinar sobre el papel de la inteligencia artificial en la justicia penal, evaluando tanto las oportunidades de prevención y eficiencia, como los riesgos ético-legales que plantea su uso.

1.3 Objetivos

El fin de este trabajo es analizar como el uso de herramientas, concretamente la Inteligencia Artificial, podrían servir como predictor del delito.

En este trabajo se buscar investigar, conocer y responder los siguientes objetivos específicos:

- Analizar el funcionamiento y fundamento de las herramientas de predicción del delito basados en algoritmos y sistemas de inteligencia artificial.
- Estudiar los principales casos internacionales de uso (PredPol y COMPAS) y sus implicaciones éticas, legales y sociales.
- Evaluar si el uso de estos sistemas sería compatible con el sistema penal español desde una perspectiva legal, constitucional y criminológica.
- Establecer una reflexión crítica sobre los beneficios, riesgos y limitaciones del uso de algoritmos y sistemas de inteligencia artificial predictiva en el ámbito penal.

1.4 Metodología

La metodología utilizada en este trabajo de investigación se basará en una investigación cualitativa de la literatura a través de una revisión de fuentes académicas y búsqueda de información relevante.

Para la realización de este trabajo, se llevó a cabo una revisión bibliográfica mediante una estrategia de búsqueda orientada a la investigación de cómo el uso de la inteligencia artificial en contextos policiales y judiciales sirve como factor predictivo del delito.

1.5 Glosario

Hotspots: Puntos de intervención

LECrim: Ley de Enjuiciamiento Criminal

LOGP: Ley Orgánica General Penitenciaria

RGPD: Reglamento General de Protección de Datos

CE: Constitución Española

CompStat: Compare Statistics

AEPD: Agencia Española de Protección de Datos

Machine Learning: Aprendizaje automático

Deep Learning: Aprendizaje profundo (combinando la IA con Machine learning)

AI Act: Reglamento de Inteligencia Artificial de la Unión Europea

Predictive Policing: Policía predictiva

Feedback Loop: Ciclo de retroalimentación

Big Data Policing: Policía basada en macrodatos

Over-policing: Sobrevigilancia

2 Estado de la cuestión

2.1 Antecedentes históricos

El origen de los sistemas predictivos aplicados al ámbito criminal no comienza con el uso de la inteligencia artificial que conocemos hoy en día, sino en los modelos actuariales que surgieron en el siglo XX. (Conceptualización de los modelos actuariales en 3.1.)

En 1928, Ernest W. Burgess desarrolla un instrumento estadístico para calcular la probabilidad de reincidencia de la actividad criminal en personas que se encontraban en libertad condicional (Burgess, 1928). Este sistema, basado en el análisis socioeconómico y antecedentes penales, son considerados los primeros intentos de predicción del comportamiento delictivo.

A lo largo del siglo XX, el avance de la criminología y el acceso a una mayor cantidad de datos y herramientas estadísticas, permitieron desarrollar métodos estadísticos más precisos para estudiar, comprender y predecir el comportamiento delictivo. Durante 1970 y 1980 se llevaron a cabo instrumentos de evaluación de riesgo como *Level of Service Inventory* (Andrews & Bonta, 1995) o *Salient Factor Score* (Hoffman & Beck, 1974), que ofrecieron la posibilidad de valorar el riesgo de reincidencia o de fracaso en libertad condicional dando asistencia en decisiones judiciales o penitenciarias. Ambos ejemplos de instrumentos fueron precursores directos de los algoritmos modernos de evaluación de riesgo como COMPAS.

En 1990, con la digitalización y automatización de los registros policiales y el desarrollo de sistemas de análisis geográficos del crimen fue cuando se produjo el primer salto hacia la predicción del delito (Perry, McInnis, Price, Smith, & Hollywood, 2013). En 1994, el Departamento de Policía de Nueva York implementó CompStat un sistema de gestión policial que utilizaba bases de datos de incidentes delictivos para generar mapas, comparar tendencias y responsabilizar a los mandos policiales por los resultados en sus distritos (Weisburd, Mastrofski, Greenspan, & Willis, 2003; Silverman, 1999).

A mediados de los 2000 aparece el programa Blue CRUSH. Un programa desarrollado en colaboración entre el Departamento de Policía de Memphis y la Universidad de Memphis. Este instrumento almacenaba grandes cantidades de datos sobre delitos para identificar zonas calientes, o más conocido, "hotspots", donde se concentraban los crímenes. Este sistema sostenía el objetivo de dirigir patrullas preventivas en los lugares con mayor probabilidad de incidencia delictiva y así optimizar los recursos policiales. (Memphis Police Department, 2006).

El auge del desarrollo tecnológico promovió una nueva generación de sistemas predictivos en la década del 2010. Un grupo de investigadores de la Universidad de California en Los Ángeles, en colaboración con el Departamento de Policía de Los Ángeles, desarrollaron el sistema PredPol, el cual comenzó a implementarse a lo largo de 2011 y 2012. Este sistema buscaba estimar la probabilidad de que un tipo específico de delito ocurriera en un área concreta y en una franja temporal determinada. Estudiaba los lugares con una mayor probabilidad delictiva (Prediction Technology Group, UCLA, 2020).

Este tipo de programas consolidaron la idea de la policía predictiva, conocida como la aplicación de algoritmos para anticipar fenómenos delictivos antes de que ocurran. A pesar de que todos estos sistemas aportaron historia y sentaron bases a los que conocemos hoy en día, también introdujeron importantes debates éticos y sociales sobre su fiabilidad y neutralidad. Uno de los sistemas más significativos fue COMPAS, analizado por ProPublica (2016), que reveló la existencia de sesgos raciales en la evaluación del riesgo de reincidencia introduciendo la duda en los límites de la objetividad algorítmica. Tras ello, el Parlamentos Europeo (2021) subrayó la importancia de garantizar transparencia, la necesidad de supervisión humana y de la protección de los derechos humanos fundamentales. Finalmente, la Agencia de Derechos Fundamentales de la Unión Europea

(FRA, 2022) destacó la continuación de sesgos algoritmos y advirtió sobre el riesgo de que los sistemas predictivos reproduzcan o amplifiquen las desigualdades ya existentes.

2.2 Avances recientes

En los últimos años el uso de algoritmos y sistemas de inteligencia artificial como herramientas predictivas ha evolucionado significativamente tanto en el ámbito académico como en el institucional. Estos avances han tenido como objetivo principal mejorar los modelos mediante técnicas de machine learning y Deep learning además de aplicar las fuentes de datos que permiten anticipar los delitos (Berk et al., 2018; Zhang et al., 2020).

Uno de los principales avances ha sido la evolución de los sistemas de predicción espacial y temporal del delito. Estos sistemas no se limitan únicamente a identificar los hotspots, sino que incorporan análisis dinámicos del comportamiento delictivo. Mohler et al. (2015) y Miró Linares -82018) presentan diferentes investigaciones mostrando como el uso avanzado de modelos espaciotemporales permite anticipar el crimen, facilitando la asignación de los recursos policiales adaptada a la realidad cambiante de los entornos. Con estas investigaciones los autores refuerzan la idea de la predicción como una herramienta de apoyo estratégico y no como un mecanismo de control reactivo.

Asimismo, en el ámbito de la justicia penal, la literatura señala que las herramientas predictivas han evolucionado positivamente en la identificación de factores de riesgo asociados a la reincidencia, a la fuga o al incumplimiento de las medidas judiciales, incrementando así su uso en las fases claves del proceso penal. A pesar de que el uso de estos sistemas muchos autores lo han visto como una oportunidad de reducir la toma de decisiones basada en la subjetividad y promover así decisiones homogéneas, otros autores no, generando un debate ético y jurídico (Monahan y Skeem, 2016).

Los avances recientes no se han limitado al ámbito técnico, sino que han fomentado un aumento de investigaciones que estudian los posibles problemas éticos y legales de estos sistemas. Tras el surgimiento, o más bien, la evolución de estos sistemas se ha intensificado el estudio de cuestiones como la transparencia algorítmica, la explicabilidad de los modelos y la necesidad de supervisión humana en la toma de decisiones automatizadas, especialmente cuando atenta a los derechos fundamentales (Burrell, 2016; Wachter, Mittelstadt y Floridi, 2017). Esta transformación evidencia una preocupación

creciente por integrar principios de justicia algorítmica en el desarrollo y uso de tecnologías predictivas.

Finalmente, todos estos avances recientes se han visto influenciados por el contexto normativo europeo, el cual ha comenzado a establecer límites y criterios para el uso de la inteligencia artificial en ámbitos sensibles. La Comisión Europea (2021) ha situado los sistemas de predicción delictiva y evaluación de riesgos en un nivel elevado de riesgo. Menciona la necesidad de garantías reforzadas, evaluación del impacto y control institucional. Este marco normativo cumple con la finalidad de orientar la investigación reciente a un enfoque más prudente, en el que la innovación tecnológica este equilibrada con la protección de principios y derechos fundamentales.

En conjunto, los avances recientes evidencian un desarrollo acelerado y multidimensional, en el que las técnicas van acompañadas de una reflexión ética y jurídica. Este doble movimiento revela que el futuro de estos sistemas no depende únicamente de su eficacia, sino también de su compatibilidad con los valores y límites.

2.3 Casos internacionales: PredPol y COMPAS como ejes principales

Los casos internacionales de herramientas algorítmicas como Predpol y COMPAS son ejemplos de cómo la inteligencia artificial se ha integrado en sistemas de justicia y seguridad como herramienta predictiva del delito y evaluación de riesgos.

Ambos casos evidencian empíricamente aplicaciones reales, efectos en decisiones operativas y judiciales y las implicaciones ético-legales (Brayne, 2017; Ferguson, 2017).

Es por ello, que su estudio se considera fundamental para comprender las consecuencias que la aplicación de la inteligencia artificial en contextos penales puede presentar, además de poder realizar una evaluación sobre la compatibilidad de esta con sistemas penales (Perry et al., 2013).

PredPol es un sistema de policía predictiva diseñado para anticipar cuándo y dónde es más probable que se cometan ciertos delitos. Este sistema emplea algoritmos destinados al análisis de datos históricos de crimen; como el tiempo, ubicación o tipo de delito, para posteriormente generar los mapas de riesgo que permitirán a las patrullas desplegarse.

Tras diferentes experimentos e investigaciones acerca de esta herramienta, Perry et al. (2013) y Mohler et al. (2015) hacen mención a dos principales oportunidades del uso del

PredPol. Este sistema permite identificar patrones espaciotemporales de la actividad delictiva, que favorecen en la planificación de patrullajes. Además de tener un enfoque en la anticipación de riesgos previos a la actividad delictiva, disminuyendo la respuesta reactiva tradicional.

Por otro lado en sus respectivos artículos, tanto Perry et al. (2013) como Mohler et al. (2015) exponen diferentes riesgos y críticas sobre este modelo. En primer lugar, la aparición de feedback loops. Los datos generados por la propia intervención policial pueden cristalizar patrones interpretados erróneamente como predictivo, pero que realmente son artefactos del sistema de vigilancia. Por otro lado, muchos de los sistemas de inteligencia artificial dificultan la auditoria pública y a rendición de cuentas por falta de datos que puedan ser contrastados o explicables.

Además, Brayne (2017) añade que PredPol tiende a dirigir la vigilancia hacia comunidades racializadas o de bajos ingresos, realizando una generalización de zonas e incrementando las desigualdades existentes. Por otro lado, argumenta sobre el bucle de retroalimentación exponiendo que este sistema utilizada datos históricos que pueden reflejar patrones de arrestos ya sesgados.

COMPAS es un software de evaluación de riesgo que se emplea en diferentes estados de EE.UU para estimar la probabilidad de reincidencia de los individuos. Es un sistema aplicado por jueces y autoridades penitenciarias como herramienta de apoyo en decisiones (Angwin et al., 1016; Ferguson, 2017).

Una de las principales oportunidades del uso del COMPAS es que es un sistema que ofrece una evaluación estandarizada de riesgos que permite complementar la información al juez (Ferguson, 2017) y que además integra múltiples factores que permiten una gestión más sistemática del riesgo, facilitando el proceso (Chouldechova, 2016).

Por otro lado, COMPAS es un sistema que presenta diferentes riesgos y críticas acerca de su manera de operar. Angwin et al. (2016) y Chouldechova (2016) mencionan que a pesar de que este sistema no recopila datos de raza directamente, realiza una correlación entre variables socioeconómicas y raza que presentan resultados desproporcionados para ciertos grupos demográficos. Ferguson (2017) habla de la falta de transparencia sobre sus procesos internos. Esta opacidad del algoritmo ha sido objeto de diferentes críticas y cuestionamientos sobre la legitimidad de este modelo predictivo. Además de haberse

visto cuestionado su uso en decisiones de libertad, argumentando que puede generar injusticias y disfrazarlas como objetividad algorítmica.

Tras realizar un breve análisis comparativo entre ambas herramientas predictivas, concluimos que PredPol se centra en predicciones espaciales, detallando datos sobre el dónde y el cuándo puede haber delitos, posibilitando así la organización de patrullas y recursos policiales. Mientras que, COMPAS se enfoca en la predicción de riesgos individuales orientados a decisiones judiciales y penitenciarias.

Ambos sistemas demuestran que los algoritmos pueden optimizar recursos y anticipar riesgos a la misma vez que pueden reproducir y amplificar desigualdades y sesgos en los datos.

2.4 Situación en España y Europa

En España el uso de inteligencia artificial en contextos de seguridad y justicia está evolucionando pero con límites y regulación. Aunque no existen sistemas de predictive policing tan estructurados como en EE.UU. si se han desarrollado herramientas que evidencian el uso de la inteligencia artificial en contextos policiales y judiciales (García Torres, 2024).

El uso de herramientas como RisCanvi o Viogén han sido empleadas como apoyo de análisis predictivo, pero es cierto que su transparencia y validación metodológica han sido objeto de cuestionamiento. El uso de Veripol, herramienta destinada para detectar denuncias falsas, tras cambios regulatorios europeos la Policía Nacional abandonó su uso (Tech y Justice, 2025).

A pesar de no existir una ley específica sobre la inteligencia artificial en ámbitos de justicia o seguridad en España, la AEPD ha tomado el mando en la supervisión del uso ético de tecnologías, sancionando el uso indebido de esta (AEDP, 2025). El Real Decreto ley 6/2023 permite el uso de herramientas de inteligencia artificial como apoyo en decisiones judiciales, mientras que, la ley orgánica 5/2024 y la ley 15/2022 imponen requisitos de transparencia y minimización de sesgos con la finalidad de que el uso de estos algoritmos no afecte a los derechos fundamentales.

En el contexto europeo existe un enfoque principal en equilibrar la innovación y uso de herramientas tecnológicas con la protección de los derechos fundamentales.

La Ley de Inteligencia Artificial de la UE clasifica los usos de la inteligencia artificial según el nivel de riesgo que suponen para los derechos y libertades fundamentales. Esta clasificación distingue entre riesgos mínimos, limitados, altos e inaceptables. Se consideran de riesgo inaceptable aquellos sistemas de inteligencia artificial que predicen el comportamiento individual y que por tanto, son prohibidos en contextos de aplicación directa en individuos (Parlamento Europeo, 2024; Le Monde, 2025).

El AI Act impone estrictas medidas de vigilancia, transparencia, evaluación de riesgo y supervisión humana para aquellos sistemas que se clasifican de alto riesgo.

Estas medidas se comenzaron a implementar a principios de 2025, todas ellas enfocadas en prohibir el uso de las herramientas de inteligencia artificial clasificadas como inaceptables; identificación biométrica en tiempo real y la social scoring. Por otro lado, las clasificadas como alto riesgo requieren evaluaciones de sesgo, auditorias y transparencias previas a su aplicación.

Aun reconociendo las críticas y limitaciones, los diversos proyectos financiados por la Comisión Europea exploran aplicaciones de análisis avanzado de datos e inteligencia artificial para apoyar análisis criminalístico, inteligencia policial o gestión de emergencias. VALCRI es un proyecto que ayuda a los investigadores a interconectar bases de datos y visualizar patrones. Es un ejemplo de tipo de proyecto que no predicen delitos de manera directa pero si ayudan en el proceso de la toma de decisiones humanas.

3 Marco teórico

3.1 Principales teorías criminológicas relacionadas con la predicción del delito: teorías actuariales, teorías ambientales, prevención situacional y teoría de la susceptibilidad diferencial

La implementación de algoritmos y sistemas de inteligencia artificial como herramientas predictivas del delito tiene su fundamento teórico en diferentes teorías en las que tratan de identificar patrones, factores de riesgo y contexto que puedan asociarse a la criminalidad. Dentro de estas teorías, distinguimos tres más influyentes: las teorías actuariales, las teorías ambientales y la prevención situacional. Todas estas han servido de base para el desarrollo de futuros modelos predictivos.

Las teorías actuariales se centran en la identificación de factores de riesgo estadísticamente asociados a la comisión de delitos, con el objetivo de estimar la probabilidad de reincidencia de un individuo o evaluar las circunstancias que suelen darse previamente a un hecho delictivo. A diferencia de la criminología clásica, orientada a explicar las causas del delito, la criminología actuarial prioriza la gestión del riesgo y la eficiencia en la toma de decisiones institucionales (Feeley y Simon, 1992). Este enfoque tiene una influencia directa en el desarrollo de herramientas algorítmicas que clasifican individuos o espacios según los diferentes niveles de riesgo. Sin embargo, desde la criminología crítica se ha indicado que esta lógica puede derivar en una justicia orientada a la administración de poblaciones y no a la responsabilidad individual, planteando así dilemas éticos y legales (Harcourt, 2007).

Las teorías ambientales desplazan el foco del individuo al entorno en el que se produce el delito. Estas teorías parten de la concepción de que la criminalidad no se distribuye de manera aleatoria sino que se concentran en espacios y momentos concretos según las oportunidades, rutinas y características del entorno. Dentro de estas se encuentra la teoría de las actividades rutinarias (Cohen y Felson. 1979), que sostiene que el delito ocurre cuando hay un infractor motivado, una víctima adecuada y la ausencia de un guardián capaz. El enfoque de esta teoría ha sido ampliamente utilizado en la predicción del delito mediante sistemas que analizan patrones espaciales y temporales para identificar los conocidos hotspots. Estos sistemas de predicción policial se apoyan en estas teorías al contemplar que la modificación de contextos y la redistribución de recursos pueden reducir las oportunidades delictivas.

Mientras que, en estrecha relación con esta última surge la prevención situacional del delito de Clarke (1995). Esta teoría propone intervenir directamente sobre las condiciones que facilitan la comisión de delitos a través de estrategias que incrementen el esfuerzo para delinquir, aumenten el riesgo percibido o reduzcan las recompensas del delito. Esta teoría ha influido notablemente en el diseño de sistemas predictivos que pretende anticipar dónde y cuándo hay mayor probabilidad de que ocurra un delito dando la posibilidad así de implementar medidas preventivas específicas. Así, estos algoritmos predictivos al facilitar intervenciones preventivas a partir de datos y análisis automatizados pueden entenderse como una extensión tecnológica de este enfoque.

Asimismo, resulta oportuno mencionar la teoría de la susceptibilidad de Belsky y O´Connor (1997). Esta teoría sostiene que las personas no responden de manera

homogénea a las influencias contextuales, sino que existen diferencias individuales en la sensibilidad a los entornos sociales. Desde esta perspectiva, ciertos individuos pueden verse más afectados por contextos adversos que otros individuos, y viceversa con los entornos favorables. Esto cuestiona los modelos predictivos rígidos que se basan exclusivamente en factores de riesgo estáticos. En el ámbito de la criminología, esta teoría introduce un enfoque relevante mencionando que el riesgo delictivo no es inmutable, sino que depende de la interacción entre características personales y condiciones sociales. Lo que plantea límites teóricos a los sistemas de predicción algorítmica del delito que están basados únicamente en datos históricos.

A pesar de que en su conjunto estas teorías criminológicas proporcionan el marco conceptual que sustenta el uso de la inteligencia artificial en la predicción del delito, su traducción a modelos algorítmicos automatizados plantea importantes restos, ya que la simplificación de fenómenos complejos y la dependencia de datos históricos pueden amplificar sesgos y desigualdades estructurales. Desde una perspectiva criminológica critica resulta imprescindible analizar como estas teorías son operacionalizadas por algoritmos y las implicaciones ético-legales en el sistema penal actual.

3.2 Conceptos clave

Para abordar el análisis de herramientas de inteligencia artificial predictivas del delito, es necesario delimitar una serie de conceptos clave que estructuran el marco teórico de este trabajo, permitiendo comprender el alcance real de estas tecnologías y evitar interpretaciones erróneas, especialmente en el ámbito penal.

En primer lugar, el concepto de inteligencia artificial refiere al conjunto de sistemas informáticos diseñados para analizar tareas que tradicionalmente requieren inteligencia humana. Dentro del ámbito penal, concretamente en la predicción del delito, la inteligencia artificial se ve reflejada a través de técnicas de machine learning. Los sistemas aprenden a partir de grandes volúmenes de datos sin necesidad de ser programados específicamente para cada situación (Russell y Norvig, 2021). Estos sistemas no únicamente ejecutan sino que ajustan sus modelos internos según los datos disponibles.

Continuando con lo anterior, el concepto de algoritmo se entiende como un conjunto de instrucciones y reglas que permiten procesar datos y obtener un resultado determinado.

En el contexto penal, estos algoritmos utilizan datos históricos para identificar patrones y niveles de riesgo que se asocian a individuos, espacios o situaciones.

Otro concepto relevante en este trabajo es el de predicción del delito. Hace referencia al uso de modelos estadísticos y computacionales para estimar la probabilidad de que se produzca un hecho delictivo en un determinado lugar, momento e individuo. Tal y como menciona Perry et al. (2013), la predicción del delito se centra en anticipar futuros escenarios con fines preventivos y de gestión de riesgo.

Finalmente, en estrecha relación con la predicción del delito, se encuentra el concepto de policía predictiva. Se entiende como el uso de herramientas algorítmicas para apoyar la toma de decisiones policiales. La policía predictiva se basa en las teorías criminológicas ambientales y actuariales, buscando optimizar la eficiencia a través del análisis de datos espaciales y temporales (Ferguson, 2017).

3.3 Fundamentos ético-legales

El uso de sistemas de inteligencia artificial como herramientas de predicción del delito plantea la necesidad de realizar un análisis de los fundamentos ético-legales que deben de regir su implementación y utilización en el ámbito penal.

Debido a que estas tecnologías son de gran repercusión en decisiones que afectan a los derechos fundamentales, su aplicación debe de evaluarse desde criterios de eficacia y eficiencia, además de someterse a los principios propios del estado de derecho y sus valores democráticos.

El articulo 10 de la CE recoge uno de los principios éticos fundamentales, el principio de dignidad humana, que actúa como eje del ordenamiento jurídico. Es por ello, que cualquier sistema tecnológico que intervenga en procesos penales debe respetar la condición del individuo como sujeto de derecho y no como objeto de un cálculo estadístico. La implementación de modelos predictivos basados en probabilidades de comportamiento puede entrar en conflicto con este principio si conllevan a decisiones que despersonalizan al individuo (Harcourt, 2007).

En relación directa con lo anterior se encuentra el principio de legalidad, el cual exige que toda actuación con consecuencias penales este en la ley y se aplique conforme a criterios claros y previsibles. La introducción de algoritmos opacos o de funcionamiento

difícilmente comprensible puede generar debate acerca de su seguridad jurídica al no ser tan fácilmente reconocibles los criterios influyentes en la toma de decisiones. Miro Llinares (2020) señala que la falta de transparencia en las herramientas algorítmicas es incompatible con nuestro modelo de justicia penal.

El articulo 14 de la CE recoge el principio de igualdad y no discriminación. Tras la revisión de la literatura, esta advierte que ciertos algoritmos predictivos pueden interferir reduciendo o amplificando desigualdades preexistentes cuando se entrenan con datos históricos sesgados, afectando de manera desproporcionada a ciertos colectivos sociales (Barocas y Selbst, 2016).

Asimismo, el articulo 24 de la CE recoge el principio de presunción de inocencia. El carácter predictivo de estos algoritmos se basa en la estimación de riesgos futuros, pudiendo entrar en conflicto con este principio si se utiliza para justificar medidas restrictivas de derechos sin la existencia de derechos probados. la penalización del riesgo en lugar de la conducta supone un desplazamiento del derecho penal del hecho a un derecho penal del autor (Feeley y Simon, 1992; Harcourt, 2007).

Desde el ámbito penal europeo, los fundamentos ético-legales han sido reforzados a partir de marcos normativos específicos en materia de inteligencia artificial y protección de datos.

El RGPD establece una serie de principios que como la licitud, la minimización de datos y la limitación, que resultan directamente aplicables con los sistemas predictivos que son utilizados en el ámbito penal. Asimismo, el Reglamento Europeo de Inteligencia Artificial ha realizado una propuesta que clasifica los sistemas de evaluación de riesgos y predicción delictiva como tecnologías de alto riesgo, obligando así a un reforzamiento de transparencias, supervisión humana y una evaluación posterior del impacto (Comisión Europea, 2021).

4 Oportunidades de los algoritmos y sistemas de IA en criminología

4.1 Optimización de los recursos policiales

La optimización de los recursos policiales refiere el proceso de intervención de una manera más efectiva y eficiente, siendo una de las oportunidades más concretas y

operativas que ofrecen los sistemas y algoritmos predictivos basados en inteligencia artificial.

A través del análisis de datos históricos y en el tiempo real, estos sistemas logran una asignación más eficiente del cuerpo, las patrullas y medios técnicos logrando una maximización de los resultados reduciendo el tiempo y coste invertido sin requerir un aumento de los recursos disponibles.

Los sistemas predictivos permiten identificar espacios y tiempos con mayor probabilidad de recurrencia de delitos, conocidos como hotspots de Sherman et al. (2014), con una mayor precisión de las herramientas tradicionales. RAND Corporation (2012) publica un informe argumentando que a pesar de no ser infalibles, hacen un uso más eficiente de los recursos para identificar estos puntos calientes. Siendo posible la anticipación de conocer el dónde y cuándo de los hechos delictivos permite incrementar la vigilancia y las patrullas dando resultados óptimos en las intervenciones.

Además, la medición y monitorización del uso de recursos facilita realizar una comparación entre la realidad y lo óptimo. Un estudio reciente expone una estrategia algorítmica basada en el uso de datos GPS, para medir dimensiones de la presencia policial en unidades microgeográficas, centrándose en la frecuencia, la duración y la intermitencia. El uso de estos datos tiene la finalidad de procesar las secuencias de los datos GPS y generar métricas precisas que permita evaluar la influencia de la presencia policial en la prevención del delito (Khalfa, Snaphaan y Hardyns, 2024).

Para demostrar que un sistema optimiza los recursos conviene hacer una medición de cinco indicadores; delitos cometidos por hora de patrulla, cambios en la tasa de delitos en micro-áreas, tiempo medio de respuesta, porcentaje de tiempo de patrullas en las zonas de riesgo y coste por delito resuelto. Indicadores que facilitan la comparación de resultados operativos a la misma vez que permiten detectar efectos de desplazamiento y orientar a los cuerpos policiales hacia áreas de mayor riesgo.

Todas estas aproximaciones dan lugar a medir tanto la presencia policial, como a realizar una distribución de recursos de forma óptima y que a su misma vez fortalezca estrategias basadas en evidencia para la prevención del delito.

4.2 Prevención del delito

Uno de los principales objetivos de los sistemas de inteligencia artificial aplicado al ámbito de la seguridad pública es la prevención del delito. Estos sistemas desarrollan la capacidad de anticipar conductas delictivas y desarrollar recursos adecuados para reducir la probabilidad de reincidencia. La conocida policía predictiva del delito no centra su trabajo únicamente en responder a los hechos ocurridos, sino adelantarse a ellos a través de técnicas analíticas y cuantitativas (Perry et al., 2013).

El fundamento teórico de la prevención mediante algoritmos se fundamenta en las teorías criminológicas de la actividad rutinaria (Cohen y Felson, 1979) y del patrón del crimen (Brantingham y Brantingham, 1993) expuestas anteriormente. Estos algoritmos, a través del análisis de numerosos datos – denuncias, llamadas, localizaciones - tratan de identificar regularidades espacio – temporales que permitan intervenir de forma temprana logrando la prevención delictual.

RAND Corporation menciona que, "los modelos predictivos permiten que los agentes tomen decisiones proactivas para prevenir delitos, no solo responder a ellos" (Perry elt al., 2013, p. 18). Esto se logra mediante la aplicación de técnicas estadísticas avanzadas – self-exciting point processes, modelos de series temporales – que se reflejan en la práctica como alertas preventivas. Los algoritmos y sistemas de predicción del delito pueden contribuir a la prevención de tres maneras. La primera, predicción de puntos calientes. Los modelos identifican los espacios urbanos con mayor probabilidad de reincidencia delictiva durante un periodo de tiempo. Este mecanismo tiene como objetivo focalizar patrullas policiales en zonas concretas consiguiendo así una reducción de hechos delictivos.

El segundo, intervención temprana focalizada. Estos sistemas permiten asociar recursos policiales a lugares, horarios o zonas de riesgo antes de que se cometa la infracción. Ferguson (2017) refiere a este mecanismo como una forma de anticipar la oportunidad delictiva.

El tercer mecanismo la prevención situacional inteligente. Los elementos predictivos tratan de modificar el entorno urbano, incluyendo los dispositivos de vigilancia, consiguiendo reducir la oportunidad de comisión.

Diferentes estudios demuestran que modelos que presentan un diseño y una implementación correcta contribuyen a reducciones medibles en los índices delictivos.

Durante un ensayo controlado aleatorio con el Departamento de Policía de Los Ángeles, se llevó a cabo una comparación de patrullajes guiados por algoritmos de predicción, dando como resultado una reducción de aproximadamente un 7% de incidentes en esta zona. Este estudio de Mohler et al. (2015) demostró que el uso de algoritmos contribuye de forma significativa en la eficiencia y predicción de delitos con recursos limitados. En reino Unido, la policía de Kent implementó un piloto de PredPol en diferentes áreas. El informe institucional de este estudió reportó una disminución de robos en las áreas donde el software se implementó en comparación con áreas en las que no. Este estudio busca confirmar que los sistemas y algoritmos de inteligencia artificial como herramientas predictivas sí pueden reducir la actividad criminal, pero esta efectividad requiere una correcta implementación y uso de las herramientas para un resultado de éxito (Kent Police, 2014). Egbert y Leese (2021) realizaron un estudio en Alemania sobre la aplicación del sistema PRECOBS como método de reducción de robos residenciales. A pesar de tener buenos resultados, no tuvieron un efecto significativo en las estadísticas de criminalidad, pero, el estudio reportó que el sistema aporto eficiencia y una mejor planificación, por lo que se considera que este estudio obtuvo un resultado positivo.

Mas allá de los resultados a nivel cuantitativo, la implementación de la inteligencia artificial en tareas preventivas supone un cambio epistemológico y operativo. A partir de la incorporación de estos sistemas, la actividad policial dejará de ser una actividad reactiva para tratar de ser preventiva a través de la gestión y priorización de escenarios de riesgo (Perry et al., 2013). Este enforque permite el inicio de la actividad política criminal focalizada en la seguridad respaldada por la evidencia, donde las decisiones se sustentan en datos y no suposiciones. Además planeta dilemas éticos y legales sobre una mala implementación de estos sistemas de inteligencia artificial.

4.3 Posibilidades en justicia penal

La aplicación de sistemas de inteligencia artificial como herramienta predictiva en el ámbito de la justicia penal manifiesta una oportunidad de modernizar los procesos judiciales, maximizar el uso de recursos y reforzar la reeducación y reinserción social de las penas. El uso de estas herramientas está permitido bajo supervisión judicial y dentro de los límites legales, únicamente cuando su principal objetivo sea optimizar la eficiencia del sistema penal alcanzando uniformidad y transparencia en el procesal en la toma de decisiones (García Torres, 2024).

En la primera etapa del proceso judicial, también conocida como fase de instrucción, el articulo 299 de la ley de enjuiciamiento criminal tiene como objetivo conocer toda la información relevante acerca del delito y sus circunstancias. Durante esta etapa, la inteligencia artificial ayudaría a la gestión inteligente de la investigación a través del análisis de pruebas digitales. La intervención de la inteligencia artificial generaría resultados exitosos a nivel general por su rapidez, además de reducir la carga de trabajo, siendo capaz de llegar a conexiones que en un investigador humano podría pasar desapercibida. Además, el articulo 118 de esta misma ley, garantiza los derechos del acusado, contemplando dentro de estos el derecho a ser informado y a tener acceso a la documentación del proceso. Durante esta fase se busca garantizar la transparencia procesal, asegurando que las partes tengan conocimiento acerca del uso de algoritmos y la influencia de estos en la decisión. Así, los sistemas de inteligencia artificial actúan como herramientas complementarias y no sustitutas del juicio humano, ajustándose al principio de exclusividad jurisdiccional que abarca el artículo 117 de la CE. En este sentido, una oportunidad de la inteligencia artificial en la justicia penal es su uso como instrumento de apoyo al juez o tribunal en tareas como valorar riesgos, establecer medidas cautelares o determinar penas. El sistema de COMPAS es un ejemplo de herramienta de evaluación de riesgo de reincidencia, fuga o quebrantamiento en la condena (Miró Linares, 2020).

En la fase de ejecución de la pena, el articulo 1 LOGP menciona que "las instituciones penitenciarias tienen como fin primordial la reeducación y la reinserción social de los sentenciados". En este contexto, el uso de la inteligencia artificial ayudaría a la personalización de programas de tratamiento, evaluar la probabilidad de reincidencia y estimar la conveniencia de medidas. Adicionalmente, los algoritmos y sistemas predictivos proporcionan utilidad en la prevención de incidentes dentro de los centros penitenciarios. Estos sistemas logran detectar patrones de violencia, ya sea autolesiones o hacia otros, incrementando los recursos evitando la reincidencia y rutinariedad de estos.

Desde un punto de vista metódico, la inteligencia artificial brinda la posibilidad de crear una justicia penal más eficiente y detectable. La automatización de las tareas y la priorización de casos conforme a criterios previamente establecidos, permitirían a los jueces y fiscales dedicar mayor tiempo a tareas de mayor valor jurídico.

A pesar de las múltiples ventajas y oportunidades que la implementación de la inteligencia artificial ofrece en este ámbito, debe respetar los principios constitucionales y procesales

básicos. Concretamente; I) Legalidad y proporcionalidad (arts. 9.3 y 25 CE), II) Tutela judicial efectiva y derecho a la defensa (art. 24 CE y 118 LECrim), III) Exclusividad jurisdiccional (art. 117 CE) y, IV) Reeducación y reinserción social (art. 25.2 y art. 1 LOGP).

4.4 Casos prácticos positivos documentados

Tras presentar los mecanismos mediante los cuales los sistemas y algoritmos de inteligencia artificial ayudan a la prevención del delito, resulta práctico analizar casos documentados que ejemplifican estas oportunidades en contextos reales.

En el apartado 4.2. se presentaron ejemplos clásicos como el estudio de Los Ángeles (Mohler et al., 2015), Kent Police (2014) y PRECOBS (Egbert y Leese, 2021). En este apartado se examinarán otros casos que también aporten evidencia sobre como la inteligencia artificial sirve de apoyo para el análisis e inteligencia criminal.

El proyecto VALCRI, "Visual Analytics for Sense-making in Criminal Intelligence Analysis" fue desarrollado entre 2014 y 2018 por la unión europea. Su objetivo principal se basó en desarrollar un entorno tecnológico avanzado que permitiera facilitar el análisis de inteligencia criminal. Para el desarrollo de este, colaboraron varias agencias policiales europeas lo que benefició a la hora de la integración de datos. Uno de los aspectos más destacados y valorados de este estudio fue su enfoque de colaboración humano-máquina, afianzando que el uso de inteligencia artificial fuera un colaborador y no un reemplazo. Asimismo, permitió la combinación de la capacidad de procesamiento de la inteligencia artificial junto con el juicio humano, maximizando así la efectividad y resultados.

Este estudio respondió con una mejora en la capacidad de análisis ya que los analistas visualizaron datos provenientes de múltiples fuentes de forma más eficiente y generando relaciones difíciles de identificar de forma manual, una evaluación positiva por parte de los usuarios y un apoyo en la toma de decisiones basada en datos focalizada en priorizar investigaciones y potenciar recursos. Demostrando que la inteligencia artificial no reemplaza al analista humano y que la integración de múltiples fuentes tiene resultados más relevantes y precisos.

Otro proyecto fueron los sistemas predictivos en España: Veripol, Riscanvi y Viogén. Se implementaron sistemas algoritmos con la finalidad de apoyar a la seguridad pública

y prevención del delito, con el objetivo principal de evaluar riesgos, potenciar los recursos disponibles y detectar posibles patrones delictivos.

Veripol (2018) es un sistema de inteligencia policial con base de inteligencia artificial que centra su trabajo en la predicción de delitos y optimización de recursos. Mientras que, Riscanvi (Alemán Aróstegi, 2023) es un sistema utilizado en el ámbito penitenciario orientado a evaluar patrones de reincidencia y desarrollar programas de intervención. Finalmente, Viogén (Ministerio del Interior, 2007) es un sistema de seguimiento de violencia de género que evalúa el riesgo de reincidencia y a partir de esto, desarrolla medidas de protección.

Estos tres sistemas presentan un enfoque muy similar al de VALCRI, donde toman la inteligencia artificial como un apoyo para poder llevar a cabo sus objetivos. En los resultados de los estudios observamos una mejora de la gestión de riesgos, tanto en la protección de víctimas como en la supervisión de las personas en situación de riesgo, así como una optimización de los recursos policiales y penitenciarios.

Adicionalmente, la inteligencia artificial aplicada en estos sistemas facilita la anticipación de comportamientos delictivos e incluso, la anticipación a estos. El análisis de datos históricos y actuales permite a los sistemas identificar patrones de riesgo que sin el uso de la inteligencia artificial no sería posible, o difícilmente. Con esto, las autoridades diseñarían estrategias de prevención fundamentadas en información objetiva.

Por ejemplo, en casos de violencia de género, el sistema podría indicar que víctimas se encuentran en situación de mayor vulnerabilidad y establecer medidas de protección más específicas. Por otro lado, en el ámbito penitenciario permite desarrollar programas de reinserción para internos con alto riesgo de reincidencia.

En conclusión, tanto los estudios mencionados, como todos los avances a lo largo de la historia, constituyen una oportunidad para incrementar la efectividad en la predicción del delito.

5 Otros posibles riesgos y límites ético-legales: una perspectiva desde la criminología

5.1 Sesgos algorítmicos y riesgo de discriminación

Los sesgos algorítmicos y de discriminación componen uno de los principales riesgos ético-legales que planeta el uso de sistemas y algoritmos de inteligencia artificial en el ámbito penal. Desde el punto de vista criminológico, el análisis de estos sesgos permite comprender como los algoritmos pueden reproducir e intensificar desigualdades estructurales existentes en la sociedad. El sesgo algorítmico se da cuando un sistema automatizado toma decisiones con un patrón de error sistemático alejándose de un resultado justo debido a los datos con los que fue entrenado. Por otro lado, el sesgo de discriminación se produce cuando estas desviaciones perjudican de forma desproporcionada a minorías generando un trato de desigualdad (Barocas y Selbst, 2016).

Son múltiples las causas por las que se producen estos sesgos y tienen relación con dinámicas sociales y criminológicas existentes. El sesgo histórico es una de las dinámicas principales que se da cuando los datos que se utilizan para programar los modelos reflejan prácticas institucionales sesgadas generando una visión distorsionada de la realidad. Este tipo de sesgo se produce cuando los algoritmos aprenden patrones basados en datos que presentan un exceso de control o vigilancia sobre determinadas zonas. Como consecuencia de este se crea una desigualdad territorial, ya que las zonas que son más vigiladas producen un mayor número de registros policiales siendo nombradas zonas de riesgo, pero sin una recogida de datos de otras zonas produciéndose el efecto feedback loop (Lum e Isac, 2016; Ensign et al., 2018).

Otro de los sesgos más destacados es el sesgo de denuncia. Este se forma cuando los algoritmos de inteligencia artificial contienen información incompleta o poco acorde con la realidad ya que únicamente recogen información acerca de los hechos denunciados sin reflejar todo aquello que no llega a conocimiento policial ocasionando que el modelo predictivo se sustente en patrones derivados de los hábitos de denuncia y no de datos totales. Este tipo de sesgo además de tener efectos técnicos, es decir, en cómo se recogen, procesan o analizan los datos, también incide en la sociedad y en el ámbito legal.

Como vemos en todos los sesgos expuestos, cuando los algoritmos se diseñan a partir de una base de datos incompleta o de información ya sesgada, estos sistemas tienden a reproducir desigualdades estructurales reforzando así los estereotipos sociales. Como

consecuencia de ello, los grupos que comúnmente han estado más expuestos al control policial – personas con bajo nivel socioeconómico o minorías étnicas – continúan siendo objeto de vigilancia. Desde la criminología este fenómeno representa una actualización del over-policing, donde la incorporación de sistemas tecnológicos intensifica la vigilancia desigual encubriéndola bajo una aparente imparcialidad y objetividad (Eubanks, 2018; Benjamin, 2019).

Eubanks (2018) sostiene que el incremento de la vigilancia algorítmica ha redefinido el control penal, desplazando la atención del individuo hacia el entorno. Este nuevo paradigma conocido como big data policing, donde el riesgo no solo se atribuye a las personas sino que también a los espacios y contextos sociales dando como resultado una criminalización indirecta de determinados territorios. De forma similar, Brayne (2017) señala que el uso creciente de dispositivos tecnológicos y grandes bases de datos en la actividad policial potencia desigualdades ya existentes dado que los datos generados y analizados están condicionados por prácticas institucionales de registro y vigilancia. Como consecuencia de ello, los algoritmos no eliminan los prejuicios humanos, sino que los institucionan logrando que los sesgos sociales se transformen en patrones automatizados de actuación.

Desde la criminología, esta automatización de la vigilancia supone un riesgo para la legitimidad del sistema penal. El uso de sistemas y algoritmos predictivos que relacionan determinadas zonas o perfiles con un mayor riesgo de delincuencia puede afectar a la ciudadanía de las instituciones disminuyendo el principio de justicia procedimental. Brayne (2017) indica que la percepción de la justicia y trato igualitario es esencial para la cooperación social y la posible eficiencia de las políticas de prevención del delito. Cuando los ciudadanos advierten de que las herramientas reproducen discriminaciones o sesgos, esta tecnología dejaría de ser útil en la mejora de la seguridad, convirtiéndose en un factor que induce la deslegitimización institucional.

La existencia de sesgos algorítmicos y de discriminación en la inteligencia artificial aplicada al ámbito penal no es un riesgo teórico, sino una realidad empíricamente comprobada. Entre los ejemplos más recurrentes en la literatura destaca el algoritmo COMPAS, considerado un caso paradigmático por postar que los sistemas automatizados pueden reproducir de manera indirecta patrones de discriminación presentes en la sociedad. Esto revela que la inteligencia artificial puede actuar como un reforzador de desigualdades, donde en la práctica, una dependencia excesiva, puede afectar a la

imparcialidad y la proporcionalidad, concretamente cuando las decisiones judiciales están basadas en puntuaciones de riesgo de sistemas no explicables. Es por ello, que el artículo 22 del RGPD 2016/679 prohíbe las decisiones con efectos jurídicos significativos que están basadas únicamente en procesos automatizados, reforzándose por el Reglamento Europeo de Inteligencia Artificial 2024/1689 que exige transparencia, supervisión humana y evaluación del impacto ético de la decisión en los algoritmos aplicados al ámbito de la justicia penal.

5.2 Opacidad de los algoritmos

Dentro de los riesgos asociados al uso de sistemas de inteligencia artificial en el ámbito penal resalta la opacidad algorítmica. Esta se entiende como la falta de transparencia acerca del funcionamiento interno de los modelos de procesamiento y análisis de datos. Esta opacidad surge como consecuencia tanto de la complejidad técnica de ciertos algoritmos como de la ausencia de mecanismos de supervisión (Burrell, 2016). Dentro del contexto penal, las decisiones pueden generar repercusiones en los derechos fundamentales, la falta de transparencia se convierte en un desafío ético y jurídico de primer orden.

Burrell (2016) defiende que la opacidad algorítmica dentro del ámbito penal se podría considerar posible en tres niveles. En primer lugar, la opacidad técnica, esta surge cuando los sistemas de predicción utilizan modelos de aprendizaje automático cuya lógica interna es incomprensible, impidiendo así comprender de forma clara el origen y motivo de las decisiones. En segundo lugar, la opacidad deliberada o también conocida como estratégica surge cuando la empresa desarrolladora de tecnologías predictivas decide de forma intencionada no revelar información relevante sobre el algoritmo penal aun conociendo su repercusión sobre los derechos fundamentales. Por último, la opacidad cognitiva se define como la falta de formación suficiente para comprender y cuestionar resultados algorítmicos.

Introducir una herramienta de incertidumbre en la toma de decisiones judiciales y policiales supone una amenaza para la legitimidad del sistema pena. Citron (2019) introduce el concepto de autoridad tecnológica explicando que cuando las partes no pueden comprender ni cuestionar el razonamiento del algoritmo se termina por vulnerar el principio de contradicción disminuyendo la confianza de imparcialidad durante el

proceso penal cuestionando el juicio crítico humano. Desde la perspectiva criminológica, la opacidad algorítmica además de repercutir a la equidad de los procedimientos altera el equilibrio entre el poder punitivo del estado y las garantías individuales.

Por consiguiente, la transparencia algorítmica además de entenderse como una exigencia técnica se entiende como una condición ética y democrática. Resulta esencial poder explicar cómo se produce una decisión automatizada para garantizar la efectividad y legitimidad del proceso judicial. En este sentido, la transparencia no debe limitarse a la expresión técnica de algoritmos sino que además debe integrarse como una cultura de supervisión y responsabilidad, en la que los jueces y fiscales comprenden los fundamentos y límites de estos sistemas tecnológicos empleados.

A través de este enfoque se otorga la posibilidad de asegurar que la inteligencia artificial actúa como una herramienta de apoyo al juicio moral y no como sustituto opaco de este.

5.3 Conflictos con otros derechos fundamentales

La implementación de sistemas de inteligencia artificial en el ámbito penal expone una serie de tensiones y conflictos con otros derechos fundamentales recogidos en la Constitución Española, en concreto, aquellos comprendidos entre los artículos 10 y 20. El articulo 10 de la CE establece que "la dignidad de la persona, los derechos inviolables que le son inherentes, el libre desarrollo de la personalidad, el respeto a la ley y a los derechos los demás son fundamento del orden político y de la paz social". Este articulo refiere la dignidad humana como eje de todo el sistema de derechos. Al introducirse algoritmos y sistemas de inteligencia artificial como herramientas de predicción delictiva se deben tener en cuenta los derechos más concretos recogidos entre el artículo 14 y el 20 de la CE, los cuales abordan la igualdad sin discriminación, el derecho a la vida e integridad, la libertad ideológica y religiosa, la libertad y seguridad personal, el derecho al honor, intimidad y propia imagen, libertad de residencia y circulación, y por último, la libertad de expresión e información.

La aplicación de la inteligencia artificial predictiva puede introducir un conflicto con los artículos expuestos. Como por ejemplo, con el artículo 14 CE. en el caso en el que un algoritmo realizara una asignación de mayores niveles de riesgo a determinados grupos por razón de raza, territorio o clase social. Asimismo, el artículo 17 CE, puede verse en una situación de conflicto cuando estos mecanismos de inteligencia artificial influyen en

decisiones de vigilancia o medidas cautelares, basadas únicamente en perfiles de riesgo automatizado sin control crítico humano. Además, el artículo 18 CE, se encuentra igualmente implicado cuando el proceso de recogida de datos, la monitorización o la predicción algorítmica incide en la intimidad de las personas sin su consentimiento.

Adicionalmente, la aplicación de la inteligencia artificial en el artículo 20 CE puede entrar en conflicto con la libertad si estos sistemas se utilizan para monitorear o predecir comportamientos delictivos de perfiles de opinión pública. La automatización del análisis de datos puede limitar implícitamente la libertad de expresión si ciertos perfiles desde el inicio son etiquetados como "alto riesgo" y por ello, objeto de mayor vigilancia.

Además, cabe mencionar que el marco constitucional exige que cualquier limitación de derechos fundamentales debe realizarse conforme a la ley, con proporcionalidad y respeto al sistema de garantías. Esto implica que los sistemas de inteligencia artificial deben de integrarse en un marco constitucional, obligando la garantía de que las decisiones automatizadas respetan la dignidad, libertad y seguridad de la persona juzgada.

5.4 Problemas de la criminología actuarial aplicada

La criminología actuarial es una rama de la criminología caracterizada por el uso de datos estadísticos, modelos probabilísticos y herramientas de evaluación estandarizadas con el fin de predecir y evaluar el riesgo de delincuencias (Feeley y Simon, 1992). Esta corriente de la criminología surgió a finales del siglo XX a raíz del giro preventivo que transformo las políticas criminales hacia una lógica más eficiente orientada al control y rehabilitación de individuos. Este enfoque criminológico ha sido utilizado en gran medida en instituciones penitenciarias, servicios de libertad vigilada y, actualmente en sistemas algorítmicos automatizados que aplican técnicas para predecir el comportamiento delictivo.

A pesar de tener el objetivo de optimizar la gestión de recursos y mejorar la toma de decisiones, su aplicación en el ámbito penal plantea una serie de problemas éticos, metodológicos y criminológicos.

Uno de los principales problemas de la criminología actuarial es su tendencia a interpretar el comportamiento humano únicamente a partir de datos numéricos, dando de lado a la complejidad de este. Harcourt (2017) señala que estos enfoques realizan una tipificación

estadística, sustituyendo el análisis de causas estructurales por una lógica puramente probabilística. Este proceso puede desencadenar una despersonalización del sujeto, comenzando como un agregado estadístico en vez de un individuo con derechos.

Otro problema significativo es la confusión entre riesgo y peligrosidad. Monahan y Skeem (2016) advierten que la mayoría de los factores de riesgo empleados por los modelos actuariales tienen un vínculo causal directo con la delincuencia y que por tanto, no siempre se podrían considerar como predictores fiables del comportamiento delictivo futuro. Elementos como el barrio de residencia, el nivel socioeconómico o los contactos previos con el sistema penal pueden establecer correlación con tasas más altas de criminalización pero esto no significa que sean indicadores de peligrosidad individual.

Un riesgo adicional es el desarrollo de justicia penal anticipatoria, donde las decisiones institucionales se basan en lo que un individuo "podría llegar a hacer" y no en lo que realmente ha hecho. Tal y como mencionan Feeley y Simon (1992) la criminología actuarial impulsa un modelo de justicia orientado al futuro, en el que la intervención penal recurre a probabilidades y categorías de riesgo en lugar de a hechos probados. Desde la perspectiva del derecho penal, este riesgo entra en conflicto con varios principios esenciales, entre ellos, la presunción de inocencia o el principio de culpabilidad. Puesto en práctica, este modelo de criminología no predice únicamente el delito, sino que abarcan diversos factores como la pobreza, la marginalidad y la exclusión social.

La combinación de la criminología con algoritmos opacos dificulta el conocimiento de variables usadas o del criterio empleado para determinar patrones de alto riesgo. La falta de explicabilidad de estos sistemas compromete la transparencia del proceso penal, además de reducir la posibilidad de que los ciudadanos cuestionen las decisiones automatizadas que podrían afectar gravemente sus derechos (Miró Llinares, 2020).

6 Discusión

A partir de la presente revisión bibliográfica, se exponen a continuación las principales reflexiones críticas derivadas del análisis realizado con los objetivos del estudio. El presente trabajo se propuso analizar el uso de algoritmos y sistemas de inteligencia artificial como herramientas predictivas, valorando sus oportunidades, riesgos y límite ético-legales desde una perspectiva criminológica y jurídico penal.

El estudio ha puesto de manifiesto que los sistemas de inteligencia artificial aplicados a la predicción del delito no constituyen una ruptura absoluta con la criminología tradicional, sino que la complementan y aportan una continuidad tecnológica de enfoques actuariales y ambientales ya existentes. Las herramientas expuestas a lo largo del trabajo, PredPol y COMPAS, se apoyan en teorías criminológicas consolidadas – teoría de la criminología actuarial, teorías de la actividad rutinaria y prevención situacional – aplicándolas en modelos automatizados capaces de procesar grandes volúmenes de datos. Frente a esta lógica actuarial, desde la teoría de la susceptibilidad diferencial se cuestiona la rigidez de las clasificaciones de riesgo, recordando que el comportamiento humano es dinámico y contextual. En consecuencia, los algoritmos son limitados para poder captar procesos de cambio, resiliencia o desistimiento delictivo. En este sentido, podemos entender la inteligencia artificial como un instrumento de gestión del riesgo, orientado a la eficiencia y a la anticipación, más que como un mecanismo explicativo.

Uno de los principales hallazgos del trabajo es que la utilidad de estos sistemas depende mayoritariamente de su concepción como herramientas de apoyo y no como sustitutos de la decisión humana. Tal y como se observa en los casos prácticos analizados, las principales ventajas de la inteligencia artificial se logran cuando esta se integra en modelos híbridos de toma de decisiones, donde le análisis automatizado complementa el juicio humano. Uno de los principales ejemplos de esta colaboración son los proyectos VALCRI o Viogén, donde evidencian que la unión humano-máquina pueden mejorar la planificación, prevención y gestión de recursos.

No obstante, el análisis crítico revela que estos beneficios conviven con riesgos estructurales relevantes, especialmente cuando los sistemas predictivos se utilizan sin garantías y sin consentimientos suficientes. Sobre todo el trabajo, ha evidenciado que los sesgos algorítmicos no son fallos excepcionales sino una consecuencia previsible de la utilización de datos históricos basados en desigualdades sociales, territoriales y raciales. Desde la criminología, esta dinámica supone una actualización del fenómeno del over-policing, en el que determinadas poblaciones y espacios quedan sometidos a una vigilancia reforzada y no por conductas individuales probadas, sino por correlaciones estadísticas. Esta lógica propone un cambio del derecho penal desde un modelo reactivo basado en hechos hacia una justicia anticipatoria, centrada en la gestión de probabilidades futuras. Como se ha explicado a lo largo del trabajo, esta tendencia entra en tensión con principios básicos.

Otro elemento clave de la discusión se sitúa en torno a la opacidad algorítmica. La falta de explicabilidad de muchos sistemas de inteligencia artificial compromete de manera grave la legitimidad del proceso penal, al dificultar el control judicial, la contradicción y la rendición de cuentas. Cuando las decisiones que afectan a decisiones atenta contra los derechos fundamentales de una persona se apoyan en modelos incompresibles para jueces o personas afectadas, produciéndose un desequilibrio entre el poder punitivo y las garantías individuales. La transparencia algorítmica no puede entenderse únicamente como una exigencia técnica, sino como una condición ética y democrática imprescindible.

Además, desde la perspectiva constitucional se evidencia que el uso de inteligencia artificial predictiva genera conflictos directos con los derechos fundamentales recogidos en la Constitución Española: Art. 14, Art 17, Art 18 y Art 24. A raíz de estos conflictos se exige un marco normativo que garantice la legalidad, proporcionalidad, supervisión humana y control institucional.

Desde el marco normativo europeo, el Reglamento de Inteligencia Artificial supone un avance significativo al clasificar estos sistemas como alto riesgo y al establecer obligaciones de evaluación, transparencia y supervisión. A pesar de ello, este trabajo expone que la regulación por sí sola no es suficiente, sino que debe de ir acompaña de una cultura institucional crítica y de una formación adecuada de los operadores.

En términos generales, la discusión permite concluir que los algoritmos y sistemas de inteligencia artificial pueden considerarse como una oportunidad relevante para la prevención del delito y la optimización de recursos, pero solo si se utilizan dentro de límites éticos, legales y criminológicos claros. Una aplicación de estos algoritmos de manera desregulada comporta riesgos significativos para la justicia, la igualdad y la legitimidad del derecho penal. La cuestión central no reside en decidir si estas tecnologías deben utilizarse, sino en definir cómo, cuándo y en qué condiciones pueden integrase sin atentar contra los principios fundamentales del Estado de Derecho.

7 Conclusiones

El presente trabajo ha tenido como objetivo analizar el uso de los algoritmos y sistemas de inteligencia artificial como herramientas predictivas del delito, valorando sus oportunidades, riesgos y límites ético-legales desde una perspectiva criminológica y

jurídico-penal. A partir de una revisión bibliográfica crítica y del análisis de casos relevantes, se puede extraer las siguientes conclusiones principales.

En primer lugar, el estudio ha puesto de manifiesto que los sistemas de inteligencia artificial aplicados a la predicción del delito no constituyen una ruptura radical con la criminología tradicional, sino que representan una continuidad tecnológica de enfoques ya existentes, especialmente de la criminología actuarial, las teorías ambientales y la prevención situacional.

En segundo lugar, se ha constatado que la inteligencia artificial puede ofrecer oportunidades relevantes en materia de prevención del delito, optimización de recursos policiales y apoyo a la toma de decisiones en el ámbito penal, siempre que su uso sea complementario y no sustitutivo del juicio humano, determinados proyectos europeos, evidencian que los mejores resultados se obtienen en modelos híbridos humano-máquina, donde la capacidad analítica del algoritmo se integra con la valoración crítica de los operadores jurídicos y policiales.

No obstante, el trabajo también concluye que el uso de estos sistemas conlleva riesgos estructurales significativos, especialmente en relación con los sesgos algorítmicos, la opacidad de los modelos y la posible vulneración de derechos fundamentales. La revisión de la literatura demuestra que los sesgos no constituyen errores excepcionales, sino una consecuencia previsible del uso de datos históricos marcados por desigualdades sociales, territoriales y raciales. Desde una perspectiva criminológica crítica, esta dinámica puede intensificar fenómenos como el over-policing y consolidar una vigilancia selectiva basada en correlaciones estadísticas y no en conductas individuales probadas.

Desde el punto de vista marco normativo, se concluye que el marco europeo reconoce el carácter sensible de los sistemas de predicción del delito y establece límites relevantes para su utilización. No obstante, la eficacia de estas garantías dependerá de su correcta aplicación práctica del desarrollo de una cultura institucional crítica y responsable.

Finalmente, desde una perspectiva criminológica, el trabajo pone de relieve que el comportamiento delictivo es dinámico y contextual, lo que limita la capacidad de los modelos predictivos para captar procesos de cambio, desistimiento o resiliencia. En consecuencia, la inteligencia artificial debe ser concebida como una herramienta auxiliar integrada dentro de un marco ético, legal y criminológico sólido. El desafío no reside en

la mera incorporación de estas tecnologías, sino en definir las condiciones bajo las cuales su uso resulta compatible con los principios fundamentales del Estado de Derecho.

En conclusión, los algoritmos y sistemas de inteligencia artificial pueden constituir una herramienta valiosa para la prevención del delito y la mejora de la eficiencia del sistema penal, pero solo dentro de límites, ético, legales y criminológicos claros. El debate no debe centrarse en si estas tecnologías deben utilizarse o no, sino en cómo, cuándo y en qué condiciones pueden integrarse sin comprometer la justicia, la igualdad y la legitimidad del derecho penal.

8 BIBLIOGRAFÍA

Larson, Julia, J. (2016, May 23). *How We Analyzed the COMPAS Recidivism Algorithm.* ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?utm_

Olenson, J., Vanbenschoten, S., Robinson, C., y Lowenkamp, C. (n.d.). *Volume 75 Number 2 Home Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments among Federal Probation Officers Background and Research Question Methods Results Discussion.* *https://www.uscourts.gov/sites/default/files/75_2_9_0.pdf?utm_*

Operation Blue C.R.U.S.H. Advances at MDP. (2025). Memphis Daily News. https://www.memphisdailynews.com/news/2006/apr/7/operation-blue-crush-advances-at-mpd/

Predictive Policing: What Is it And Should It Be Used In 2020? (2020, June 23). UC Institute for Prediction Technology. https://predictiontechnology.ucla.edu/predictive-policing-what-is-it-and-should-it-be-used-in-2020/?utm_

Andrews, D. A., y Bonta, J. (1995). *The Level of Service Inventory-Revised.* Multi-Health Systems.

Burguess, E. W. (1928). *Factors determining success or failure on parole*. The University of Chicago Press.

European Parliament. (2021*). Ensuring transparency, oversight, and accountability*. https://www.europarl.europa.eu

FRA- European Union Agency for Fundamental Rights. (2022). *Bias in algorithms - Artificial intelligence and discrimination.* Publications Office of the European Union.

Hoffman, P. B., y Beck, J. L. (1974). *Salient Factor Score: A procedure for evaluating parole prognosis. Journal of Criminal Justice, 2*(3), 195-206.

Memphis Police Department. (2006). *Blue CRUSH: Crime Reduction Utilizing Statistical History*. City of Memphis.

Perry, W. L., Mclnnis, B., Price, C. C., Smith, S. C., y Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation.

ProPublica. (2016). *Machine Bias: There´s software used across the country to predict future criminals. And it´s biased against blacks.* https://www.propublica.org

Silverman, E. (1999). *NYPD Battles Crime: Innovative Strategies in Policing.* Northeastern University Press.

Weisburd, D., Mastrofski, S. D., Greenspan, R., y Willis, J. J. (2003*). CompStat and organizational change: A national assessment.* U. S. Department of Justice.

Kent Police. (2014). *PredPol operational review – initial findings*. https://www.statewatch.org/media/documents/docbin/uk-2013-11-kent-police-pp-report.pdf

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., y Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association, 110* (512), 1399 – 1411.

Eubanks, V. (2018). *Automating Inequality: How high-teach tools profile, police, and punish the poor*. St. Martin's Press.

Brayne, S. (2017). Big data surveillance: the case of policing. *American Sociological Review, 82*(5), 977-1008. https://doi.org/10.1177/0003122417725865

Feeley, M., y Simon, J. (1992). The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology, 30*(4), 449-474.

Harcourt, B. (2017). *Against Prediction: profiling, Policing, and Punishing in an Actuarial Age.* University of Chicago Press.

Monahan, J., y Skeem, J. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12,* 489-513.

Miró Llinares, F. (2020). *Ética, transparencia y control de algoritmos en el proceso penal.* Revista de Derecho Penal Contemporáneo, 23(1), 1-19.

Brantingham, P. J., Valasik, M., y Mohler, G. O. (2018*). Does predictive policing lead to biased arrests? Results from a randomized controlled trial.* Statistics and Puublic Policy, 5(1), 1-6.

Burrell, J. (2016). *How the machine "thinks": Understanding opacity in machine learning algorithms.* Big Data & Society, 3(19), 1-12.

Cohen, L., y Felson, m. (1979). *Social change and crime rate trends: a routine activity approach.* American School Review, 44(4), 588-608.

Brantingham, P. J., y Brantingham, P. L. (1993). Environment, routine and situation: Toward pattern theory of crime. In R. V. Clarke y M. Felson (Eds.), Routine Activity and Rational Choice.

Citron, D. K. (2019). *Technological due process*. Washington University Law Review, 85(6), 1249-1313.

Egbert, S., y Leese, M. (2021). Criminal futures: Predictive policing and everyday police work in Germany. Science, Technology, and Human Values, 46(2), 254-279.

Ferguson, A. G. (2017). *The rise of big data policing: Survelliance, race, and the future of law enforcement.* Ney York University Press.

Khalfa, J., Snaphaan, T., y Hardyns, W. (2024). *Measuring police presence using GPS data: A micro-geographic approach to proactive policing.* Journal of Quantitative Criminology.

Sherman, L. W., Neyroud, P., y Neyroud, E. (2014). *The Cambridge Crime Harm Index: Measuring total harm from crime based on sentencing guidelines*. University of Cambridge Press.

Alemán Aróstegi, M. (2023). *Evaluación del programa RisCanvi en instrucciones penitenciarias.* Revista Española de Investigación Criminológica, 21(1), 1-28.

Miró Llinares, F. (2020). Ética, transparencia y control de algoritmos en el proceso penal. *InDret, 4*, 1–30.

Harcourt, B. (2007). *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. University of Chicago Press.

Feeley, M., & Simon, J. (1992). The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology, 30*(4), 449–474.

Comisión Europea. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation.

Clarke, R. V. (1995). Situational crime prevention. *Crime and Justice, 19*, 91–150.

Statewatch (2022). *Civil society calls on the EU to prohibit predictive and profiling AI systems in law enforcement and criminal justice*.

García Torres, M. L. (2024). *La inteligencia artificial predictiva al servicio de la prevención e investigación del delito y del proceso penal*. *Ciencia Policial*, 183, 91–132.

Agencia Española de Protección de Datos (2025). *España lidera en sanciones por protección de datos con casi 19 000 reclamaciones anuales*. El País.

Artificial Intelligence Act: MEPs adopt landmark law (2024). Parlamento Europeo.

España | AI rolled out in justice system in digitalisation strategy (2025). Tech & Justice (Oxford).

Chouldechova, A. (2017). *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments.*

Brayne, S., Rosenblat, A., & Boyd, D. (2015). *Predictive Policing and Civil Rights Primer.* Data & Civil Rights (incluye perspectivas críticas).

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99.

Monahan, J., & Skeem, J. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12*, 489–513.

Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). *Comparison of machine learning algorithms for predicting crime hotspots*. *IEEE Access, 8*, 181302–181310. https://doi.org/10.1109/ACCESS.2020.3028420

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2018). *Fairness in criminal justice risk assessments: The state of the art*. Sociological Methods & Research, 50(1), 3–44. https://doi.org/10.1177/0049124118782533