La genética del algoritmo en la política criminal: herencia, sesgo y racionalidad punitiva en la era de la IA

The Genetics of the Algorithm in Criminal Policy: Inheritance, Bias, and Punitive Rationality in the Age of AI

Oswaldo R. Aguilar Rivera

Doctor en Derecho por el Instituto de Investigaciones Jurídicas, UNAM
Docente en el Instituto Nacional de Ciencias Penales
Correo electrónico: oswaldoaguilarrivera@gmail.com
ORCID: https://orcid.org/0009-0009-6339-8734

La genética del algoritmo en la política criminal: herencia, sesgo y racionalidad punitiva en la era de la IA

The Genetics of the Algorithm in Criminal Policy: Inheritance, Bias, and Punitive Rationality in the Age of AI

Oswaldo R. Aguilar Rivera
Instituto Nacional de Ciencias Penales

Recepción: 12/03/2026 **Aceptación:** 14/04/2026

DOI: https://doi.org/10.57042/rmcp.v9i29.1044

Resumen

Este artículo propone la noción de genética del algoritmo para analizar críticamente el uso de inteligencia artificial (IA) en la política criminal. Sostiene que estos sistemas incorporan herencias de datos, categorías y prioridades institucionales que condicionan su funcionamiento y pueden reforzar selectividad penal y racionalidad punitiva. A partir de ello, se plantean límites democráticos y garantistas para su utilización.

Palabras clave

Genética del algoritmo, política criminal, inteligencia artificial, selectividad penal, racionalidad punitiva.

Abstract

This article proposes the notion of the genetics of the algorithm as an analytical category for critically examining the use of ai in criminal policy. It argues that these systems incorporate inherited data structures, institutional categories, and policy priorities that shape their operation and may reinforce penal selectivity and punitive rationality. On that basis, the article advances democratic and due-process-based limits for their use.

Keywords

Genetics of the algorithm, criminal justice policy, artificial intelligence, penal selectivity, punitive rationality.

Sumario

I. Introducción. II. Política criminal y tecnologías de anticipación. III. Hacia una definición de genética del algoritmo. IV. La herencia del algoritmo. V. Efectos de la herencia algorítmica: sesgo, selectividad y racionalidad punitiva. VI. Límites democráticos y garantistas del uso de la IA en la política criminal. VII. Referencias.

I. Introducción

Con frecuencia, la incorporación de sistemas de IA en el ámbito penal se ha promovido en nombre de la innovación, la eficiencia y la objetividad. En ese registro, la tecnología suele aparecer como una herramienta capaz de optimizar decisiones, mejorar la asignación de recursos y fortalecer la capacidad institucional de prevenir, investigar y gestionar riesgos. Sin embargo, en materia penal esa narrativa requiere una revisión crítica más exigente, pues las herramientas algorítmicas no operan sobre un terreno

neutro, sino en un espacio históricamente atravesado por vigilancia diferencial, selectividad, desigualdad estructural y expansión de las capacidades de control estatal.

En ese contexto, el presente trabajo parte de una inquietud central: los problemas que suscita la inteligencia artificial en la política criminal no pueden explicarse sólo a partir de sus resultados visibles ni reducirse a cuestiones de precisión técnica, opacidad o error operativo. Antes bien, requieren una categoría que permita examinar las condiciones históricas, institucionales y normativas que configuren desde el origen el funcionamiento de estos sistemas. A partir de esa premisa, este artículo propone el término genética del algoritmo como herramienta analítica para comprender que los sistemas de inteligencia artificial aplicados al campo penal incorporan una herencia de datos, categorías, prioridades institucionales y márgenes de tolerancia al error que condicionan su diseño, su operación y sus efectos.

La hipótesis que orienta estas páginas es que la inteligencia artificial no constituye, en el ámbito penal, una simple innovación instrumental, sino un dispositivo técnico-político cuya inserción puede reforzar formas previas de selectividad, clasificación desigual y racionalidad punitiva, particularmente cuando su uso se legitima a partir de criterios de eficiencia desvinculados de exigencias democráticas y garantistas. Desde esta perspectiva, el artículo se propone, en primer término, situar a la inteligencia artificial en el marco de una política criminal cada vez más orientada a la anticipación y la gestión del riesgo; posteriormente, construir el concepto de genética del algoritmo y examinar las distintas capas de herencia que lo integran; finalmente, analizar sus efectos sobre la selectividad penal y plantear algunos límites democráticos y garantistas para

su utilización en el control penal. El objetivo es aportar una clave de lectura que permita someter a la IA y su uso en la política criminal a un examen crítico de procedencia, legitimidad y compatibilidad con los derechos fundamentales.

II. Política criminal y tecnologías de anticipación

La incorporación de sistemas de inteligencia artificial en el ámbito penal no representa una ruptura absoluta con las formas tradicionales de control social, sino más bien la profundización de una transformación que desde hace tiempo reconfigura la política criminal. En efecto, una de las mutaciones más significativas del pensamiento penal contemporáneo ha consistido en el desplazamiento paulatino de un modelo predominantemente reactivo, centrado en la respuesta frente al hecho consumado, hacia un modelo cada vez más orientado a la anticipación, la prevención y la administración diferencial del riesgo (Feeley y Simon, 1992). Esta transición no es menor, porque implica una modificación en la racionalidad desde la cual el Estado observa, clasifica y gestiona a las personas, los territorios y las conductas consideradas problemáticas.

La política criminal clásica, al menos en su autocomprensión liberal, tendía a justificarse como una respuesta excepcional frente a la lesión o puesta en peligro de bienes jurídicos previamente definidos por la ley. A partir de ese paradigma, el derecho penal aparecía como una reacción institucional ante hechos determinados, probados y atribuibles a sujetos concretos. Sin embargo, el desarrollo de nuevas estrategias de seguridad, la expansión del discurso de la prevención, la consolidación de enfoques actuariales e incluso el populismo punitivo han alterado esa gramática.

La preocupación que antes se enfocaba en el hecho delictivo realizado, ahora lo hace en la identificación temprana de probabilidades, correlaciones, patrones y amenazas potenciales. Esta mutación transita el eje de la imputación retrospectiva hacia la sospecha prospectiva.

En este contexto, la inteligencia artificial no inaugura por sí sola una nueva racionalidad, pero sí potencia de manera extraordinaria tendencias que ya se encontraban en desarrollo. Su promesa institucional radica precisamente en ofrecer una capacidad inédita para procesar grandes volúmenes de información a través de la *big data*, identificar regularidades, detectar anomalías y producir clasificaciones automatizadas que supuestamente permitirían intervenir antes de que el daño se materialice. De este modo, la IA se presenta como una aliada natural de una política criminal cada vez más interesada en la previsión, la priorización y la administración eficiente del riesgo. No obstante, esa afinidad entre esta tecnología y política criminal anticipatoria no debe leerse como un simple avance técnico, sino como la convergencia entre una herramienta poderosa y una racionalidad institucional previamente inclinada hacia el control preventivo.

La vigilancia ocupa un lugar central en esta transformación. En la medida en que la política criminal se orienta a anticipar escenarios de riesgo, necesita ampliar su capacidad de observación sobre los comportamientos, desplazamientos, vínculos y hábitos de las personas, de acuerdo con su genética, a lo cual, en reflexiones previas, he denominado

genética de la vigilancia.[1] La vigilancia era concebida como un instrumento auxiliar de investigación, y ahora es un acto protagonista de la condición estructural de la prevención. A partir de ahí, la recolección masiva de datos, el monitoreo constante, la interconexión de bases de información y el análisis automatizado adquieren una relevancia estratégica. Por eso, los sistemas algorítmicos encuentran un terreno fértil en instituciones que ya habían asumido que el conocimiento anticipado del riesgo requiere observación intensiva y permanente.

Ahora bien, esa vigilancia ampliada no opera de forma neutra. Para ser funcional a la política criminal, la información debe traducirse en categorías operativas. Los sistemas de IA parten de observar, recolectar y clasificar, muchas veces a través del ojo humano mediante el *crowdsourcing*[2] (Crawford, 2021). Identifican perfiles, jerarquizan prioridades, distinguen niveles de amenaza, marcan desviaciones y ordenan poblaciones conforme a parámetros previamente definidos. En este sentido, la clasificación es una operación política antes que meramente técnica, porque determina qué características serán consideradas relevantes, qué indicadores se asociarán con el peligro y qué formas de comportamiento serán interpretadas como señales de alerta. El problema no reside únicamente en que los algoritmos pue-

1 Transmisión y perfeccionamiento de las estructuras y los patrones de la vigilancia con el fin de perpetuar los sistemas de recolección y procesamiento de datos.

2 Microtrabajadores que efectúan las repetitivas tareas digitales que subyacen a los sistemas de IA (como etiquetar miles de horas de datos de entrenamiento y revisar contenidos sospechosos o dañinos).

dan equivocarse, sino en que su propia lógica clasificatoria condensa decisiones previas sobre lo normal, lo sospechoso, lo tolerable y lo intolerable.

Esta capacidad de predicción y clasificación transforma profundamente la relación entre conocimiento y poder punitivo. En el modelo tradicional, el poder penal se activaba principalmente después de la infracción. En el paradigma anticipatorio, en cambio, el conocimiento probabilístico habilita intervenciones previas, selectivas y desiguales. Allí donde un sistema identifica concentraciones de riesgo, sujetos prioritarios o territorios críticos, el aparato de control tiende a intensificarse. Entonces, la predicción ya no sólo describe una realidad, sino que contribuye a producirla, porque orienta la distribución del poder punitivo y consolida focos de control diferencial.

Desde esta perspectiva, las tecnologías de anticipación participan en una expansión funcional del poder penal. No siempre lo hacen mediante la creación formal de nuevos delitos o el endurecimiento explícito de penas, sino a través de mecanismos más sutiles de intensificación del control. El poder punitivo se expande cuando aumenta su capacidad para observar, clasificar, seguir, perfilar y priorizar intervenciones sobre sujetos y poblaciones específicas. También se expande cuando el umbral de sospecha práctica se reduce por efecto de herramientas que prometen detectar riesgos antes de que se manifiesten en actos jurídicamente relevantes. En ese sentido, la IA puede operar como multiplicadora del alcance estatal, al dotar a las instituciones de instrumentos más penetrantes para intervenir en fases tempranas, difusas o incluso inciertas del comportamiento humano.

Lo anterior obliga a tomar distancia frente a las narrativas que presentan estas herramientas como simples in-

novaciones de eficiencia administrativa. La eficiencia no es una categoría inocente cuando se inserta en el campo penal. Una política criminal más eficiente en la identificación, seguimiento y clasificación de riesgos puede ser, al mismo tiempo, una política criminal más invasiva, más selectiva y más difícil de controlar democráticamente.

Por ello, la relación entre política criminal e inteligencia artificial debe ser comprendida en clave histórica e institucional. ¿La IA aterriza en un vacío normativo y en una estructura neutral de administración de justicia? ¿O se inserta en instituciones atravesadas por inercias selectivas, desigualdades estructurales, prioridades de control y marcos de interpretación sobre el riesgo social? La novedad no es menor ni sencilla porque no radica en haber inventado la lógica anticipatoria, sino en haberla dotado de una capacidad de cálculo, segmentación y proyección sin precedentes. Dicho de otro modo, la inteligencia artificial intensifica tendencias previas: acelera la conversión del dato en sospecha (que políticamente se describe como tendencia), amplifica la vigilancia como método ordinario de prevención y robustece una política criminal que aspira a intervenir antes.

En consecuencia, cualquier aproximación crítica a los sistemas algorítmicos en el ámbito penal debe partir de que éstos se insertan en una política criminal ya orientada a la anticipación, la clasificación y la gestión del riesgo.

III. Hacia una definición de genética del algoritmo

Una de las dificultades más frecuentes en la discusión contemporánea sobre la IA y el sistema penal consiste en que gran parte de la crítica se ha concentrado en categorías que, aunque indispensables, resultan todavía insuficientes

para explicar la profundidad del problema. Nociones como sesgo algorítmico u opacidad permiten identificar efectos relevantes, pero no alcanzan por sí solas a explicar de dónde provienen los criterios que orientan al sistema ni cómo se incorporan a su estructura (O´Neil, 2016). El riesgo de limitar la crítica a esas categorías consiste en permanecer en el nivel de los síntomas sin alcanzar todavía el plano de la constitución del problema.

En efecto, cuando un sistema algorítmico reproduce discriminaciones o intensifica patrones de selectividad penal, la explicación no puede agotarse en afirmar que fue entrenado con datos sesgados o que su lógica interna no es del todo transparente. Esas afirmaciones son correctas, pero resultan todavía parciales. Lo verdaderamente relevante es advertir que, antes de que el sistema procese o clasifique, ya existe una serie de decisiones sobre los datos que incorporará, las categorías que organizarán su operación y los fines institucionales que justifican su implementación. En otras palabras, el problema del algoritmo comienza antes de su funcionamiento visible ya que radica en su herencia. Por ello, para comprender críticamente su inserción en la política criminal, hace falta una categoría que permita examinar su constitución estructural y no sólo sus manifestaciones externas.

En ese punto cobra sentido la noción de genética del algoritmo. La expresión, más que biologizar la tecnología o sugerir una analogía naturalista rígida, busca ofrecer una herramienta conceptual para pensar los elementos originarios, heredados y estructurales que determinan la orientación de un sistema algorítmico. Hablar de genética, en este contexto, significa desplazar la atención desde el resultado inmediato hacia la composición profunda del dispositivo; los datos que lo alimentan, las categorías que organizan

su operación, los fines institucionales que lo orientan y las lógicas de clasificación que condicionan sus efectos. La genética del algoritmo remite, así, a aquello que antecede a su funcionamiento visible y que, sin embargo, delimita de manera decisiva lo que puede ver, inferir y jerarquizar.

Antes de avanzar, conviene precisar qué se entiende aquí por algoritmo. En su acepción básica, se trata de un conjunto ordenado y finito de operaciones orientado a resolver un problema. Sin embargo, en el contexto contemporáneo esa definición resulta insuficiente si no se advierte que los algoritmos operan dentro de entramados sociales, institucionales y éticos específicos. En el ámbito penal, por ello, no deben ser vistos sólo como secuencias técnicas de procesamiento, sino como dispositivos que traducen decisiones humanas en criterios de clasificación, predicción y priorización con efectos concretos sobre derechos y libertades.

Esta precisión pretende ser útil en el ámbito penal, para el cual los algoritmos, lejos de operar en un vacío técnico, lo hacen en un terreno históricamente atravesado por prácticas de vigilancia, clasificación y control. Por ello, la genética del algoritmo no debe entenderse únicamente como una cuestión informática, sino como una estructura técnico-política. Lo que el algoritmo "es" no depende sólo de su código, sino también de las decisiones institucionales que le dan forma, de las racionalidades de control que lo justifican y de los contextos de desigualdad sobre los que interviene. Por tanto, su genética no se reduce a la dimensión computacional del sistema, sino que incluye la procedencia social y normativa de sus componentes. Lo importante no es únicamente cómo funciona, sino desde qué presupuestos ha sido construido para funcionar de ese modo.

A partir de ello, puede proponerse la siguiente definición de genética del algoritmo:

> [...] la transmisión, reproducción y perfeccionamiento de las estructuras internas, los patrones de cálculo y las lógicas de decisión que configuran los sistemas algorítmicos, orientados a incrementar su capacidad de procesamiento, optimización y adaptación, mediante ciclos continuos de entrenamiento, retroalimentación y ajuste que consolidan su expansión funcional y su integración progresiva en entornos institucionales y sociales.

Particularmente, este concepto debe insertarse en la discusión de contextos de vigilancia, clasificación y control penal. En esta línea, se ha mostrado que la vigilancia no debe entenderse sólo como observación ocasional o simple supervisión, sino como una atención focalizada, sistemática y rutinaria sobre detalles personales con fines de influencia, gestión, protección o dirección. A partir de esta premisa, la vigilancia contemporánea se convierte en una práctica estructurante de producción de conocimiento y ejercicio de poder sobre las personas y las poblaciones (Lyon, 2007). Tales elementos comprenden, entre otros, la selección de los datos de entrada, las categorías mediante las cuales esos datos son organizados, los objetivos institucionales perseguidos, los parámetros de clasificación, los márgenes de error aceptados y las lógicas político-criminales que subyacen a su implementación. Desde esta perspectiva, el algoritmo deja de aparecer como un mero instrumento de procesamiento para convertirse en una condensación de decisiones previas que delimitan lo que puede ver, lo que puede inferir, lo que puede jerarquizar y, en última instancia, sobre quiénes puede proyectar sus efectos.

La utilidad de esta noción radica también en que permite distinguir entre el funcionamiento técnico del algoritmo y su estructura genética. El primero remite al modo en que el sistema procesa datos reconoce patrones o produce clasificaciones una vez puesto en marcha; la segunda, en cambio, alude a las condiciones previas que orientan su diseño y delimitan sus efectos. Esta diferencia es crucial, porque un algoritmo puede operar correctamente desde el punto de vista técnico y, sin embargo, resultar normativamente problemático por las bases sobre las que ha sido construido. La crítica, por tanto, no debe agotarse en evaluar si el sistema funciona, sino en examinar la legitimidad de aquello que lo hace funcionar de ese modo.

Además, esta distinción permite evitar un equívoco frecuente en el entusiasmo tecnocrático: asumir que una mayor sofisticación del sistema basta para corregir los problemas que plantea su uso en el ámbito penal. No necesariamente es así. Un modelo más refinado puede reproducir las mismas prioridades institucionales, las mismas categorías de clasificación y las mismas asimetrías heredadas. Por ello, el examen crítico del algoritmo no debe centrarse sólo en su rendimiento, sino también en la procedencia y orientación de sus componentes.

Otro aspecto decisivo de esta categoría es que permite comprender que la herencia algorítmica no es exclusivamente un problema de datos. La genética del algoritmo incorpora también categorías, finalidades y lenguajes institucionales que convierten ciertos fenómenos en objetos de vigilancia y clasificación. En el campo penal, esto significa que el sistema no hereda únicamente información, sino también formas históricas de mirar, ordenar y administrar el desorden social e implícitamente, situar a ciertos sujetos dentro de marcos previos de sospecha, riesgo y control (Arteaga, 2017).

Desde esta óptica, la genética del algoritmo ofrece una ventaja analítica importante para el análisis penal y criminológico porque permite conectar la discusión tecnológica con problemas clásicos de la política criminal, la criminología y los derechos humanos. En lugar de concebir a la inteligencia artificial como una caja negra cuyas salidas deben corregirse externamente, esta categoría permite reconstruirla como un artefacto cuya configuración participa ya de las lógicas institucionales de clasificación y control.

De este modo, la genética del algoritmo permite articular, en una sola clave de análisis, la dimensión criminológica, político-criminal y jurídico-garantista del problema. Para ello, descarta el análisis desde la cuestión técnica o del reduccionismo de una denuncia abstracta de discriminación digital y plantea un análisis de una forma específica de organización del control que exige ser examinada a la luz de sus presupuestos, sus efectos y sus límites. Desde una perspectiva cercana al garantismo penal de Luigi Ferrajoli, no basta con que una herramienta sea funcional para la persecución o prevención del delito; es indispensable que su diseño, implementación y efectos puedan someterse a criterios de legalidad, control, racionalidad y protección de los derechos fundamentales (Ferrajoli, 1995). En suma, la genética del algoritmo permite articular estos tres planos sin reducir el fenómeno a una mera cuestión técnica ni a una simple denuncia abstracta de la discriminación digital.

Asumir esta categoría implica reconocer que los sistemas de inteligencia artificial utilizados en el ámbito penal deben ser evaluados a partir de las condiciones históricas, institucionales y normativas que configuran su funcionamiento. En esa medida, la genética del algoritmo permite desplazar la atención desde la superficie operativa del sistema hacia la estructura que orienta sus efectos. Desde esa

base, resulta posible examinar, en el apartado siguiente, las distintas formas de herencia que lo atraviesan. En suma, la noción de genética del algoritmo permite dar un paso conceptual que resulta indispensable para este trabajo: permite trascender la crítica centrada en fallas visibles a una crítica orientada a las condiciones de producción del sistema; pasar de la pregunta por el error a la pregunta por la herencia; pasar de la superficie operativa del algoritmo a la estructura político-institucional que lo constituye. A partir de esta clave, la IA deja debe revelarse como un dispositivo históricamente situado, normativamente cargado y estructuralmente condicionado por la política criminal en la que se inserta. Ese desplazamiento analítico será decisivo para comprender, en los apartados siguientes, cómo la herencia del algoritmo puede traducirse en sesgo, selectividad y racionalidad punitiva.

IV. La herencia del algoritmo

La pregunta decisiva en este punto consiste en identificar qué heredan los sistemas algorítmicos cuando se insertan en el ámbito penal. La herencia del algoritmo remite, precisamente, a esa memoria previa que condiciona su funcionamiento y delimita sus efectos: desde la aparición de la escritura, por ejemplo; existen indicios de que la civilización sumeria recurrió a sistemas matemáticos para resolver problemas aritméticos de gran escala, como la distribución proporcional de granos, lo que revela que la formulación de secuencias operativas para ordenar, calcular y administrar recursos no es enteramente nueva, aunque hoy adopte formas tecnológicamente más sofisticadas (Chabert, 1994). Antes de clasificar, priorizar o predecir, estos sistemas ya

han sido configurados a partir de determinados insumos, categorías y finalidades institucionales.

La primera dimensión de esa herencia está constituida por los datos. En el ámbito penal, éstos no pueden ser entendidos como reflejos neutros de la realidad, sino como registros producidos a partir de prácticas institucionales de observación, selección y vigilancia desiguales. Detenciones, denuncias, patrullajes, investigaciones o bases administrativas expresan no sólo hechos, sino también modos previos de intervención estatal sobre determinados sujetos, territorios y conductas.

Cuando un algoritmo es entrenado con datos históricos, recibe más que información sobre hechos pasados: recibe la huella de decisiones previas que determinaron qué fue registrado, sobre quién se concentró la observación institucional y en qué espacios se produjo una mayor densidad de intervención. En ese sentido, el sistema aprende de eventos y de la forma en que las instituciones han percibido y administrado previamente el desorden social.

Pero la herencia del algoritmo no se agota en los datos. Una segunda dimensión fundamental se encuentra en las categorías institucionales con las que esos datos son organizados y traducidos en objetos de cálculo. Para que un sistema de inteligencia artificial opere, no basta con alimentarlo de registros; es necesario definir qué variables importan, qué indicadores resultan relevantes y qué clasificaciones serán útiles para la intervención. Así, el algoritmo obtiene una gramática institucional que convierte ciertos fenómenos en riesgo, sospecha o prioridad, y que orienta desde el origen la manera en que el sistema observará y jerarquizará la realidad penal.

Además, en cuanto al riesgo, hoy la etiqueta de "peligrosidad" sale de los parámetros de protección de derechos

humanos y entra en la categoría de estigmatización y criminalización. Algo semejante ocurre con categorías como "zona criminógena", "perfil de alto riesgo" o "patrón de comportamiento sospechoso", que describen un modo de ver a ciertas personas y territorios desde el signo de la amenaza. El algoritmo, al trabajar con esas categorías, no se limita a reproducirlas mecánicamente: las estabiliza, las proyecta y las reviste de una apariencia renovada de objetividad computacional.

Existe, además, una tercera capa de herencia: la de las prioridades político-criminales. Todo sistema algorítmico implementado en el ámbito penal responde a una decisión institucional previa acerca de para qué se le considera útil. Esa orientación no es técnica, sino política, porque expresa qué riesgos se pretende anticipar, qué conductas se consideran prioritarias y qué formas de intervención se buscan fortalecer. De este modo, el algoritmo recibe, además de información y categorías, finalidades institucionales que delimitan el horizonte dentro del cual adquiere sentido.

Desde esta perspectiva, el algoritmo adquiere la lógica de las prioridades institucionales que justifican su existencia. Lo que aprende, clasifica o proyecta depende en gran medida de los objetivos que orientaron su diseño. Por ello, la herencia algorítmica no consiste únicamente en contenidos informacionales, sino también en finalidades político-criminales que condicionan su operación desde el origen. Si una política criminal privilegia la prevención situacional (Clark, 1997), el algoritmo tenderá a ser configurado para identificar espacios de concentración delictiva y orientar vigilancia intensiva. Si privilegia la neutralización de sujetos considerados peligrosos, el sistema tenderá a organizar la información en torno a perfiles individuales o trayectorias de riesgo e incluso grupos históricamente ex-

cluidos. Si se orienta por una lógica de eficiencia administrativa, el algoritmo será valorado por su capacidad para jerarquizar casos, reducir tiempos o asignar recursos de forma óptima.

A estas dimensiones debe añadirse otra igualmente importante: la herencia de las lógicas históricas de vigilancia. En este sentido, el panóptico de Bentham constituye un antecedente fundamental dentro de lo que el autor denomina la genealogía de la vigilancia.[3] En el campo penal, las instituciones no observan de manera homogénea al conjunto social, sino que históricamente concentran su atención sobre ciertos territorios, grupos y formas de vida. Estas asimetrías no desaparecen con la introducción de la inteligencia artificial; por el contrario, pueden sedimentarse en los sistemas algorítmicos cuando éstos se alimentan de registros producidos precisamente desde esa vigilancia desigual.

La herencia del algoritmo incluye, así, una memoria territorial y social del control. Lo que significa que el sistema puede reorganizar como indicadores de riesgo aquellas zonas y poblaciones que históricamente han sido más expuestas al registro institucional. De este modo, la inteligencia artificial no sólo procesa información pasada, sino que puede consolidar distribuciones heredadas de sospecha y vigilancia desigual.

3 Origen y desarrollo histórico de la vigilancia que tiene por objeto conocerlo todo (hábitos, patrones, emociones, orientaciones, ubicaciones), a través de cualquier medio de recolección de datos, y que busca perpetuarse con modelos cada vez más eficientes e imperceptibles para el ser humano.

Hay todavía una última dimensión de la herencia que merece atención: la de los márgenes de error socialmente tolerados. Todo sistema algorítmico incorpora una decisión, explícita o implícita, acerca de quiénes habrán de soportar preferentemente los costos del error. En materia penal, esta cuestión es especialmente delicada, porque detrás de esos márgenes se define cuánto control injustificado, cuánta vigilancia innecesaria o cuánta sospecha anticipada está dispuesta a tolerar una institución en nombre de la seguridad.

Esta dimensión muestra con claridad que la herencia del algoritmo, más allá de insumos informacionales, requiere decisiones normativas acerca del balance entre libertad, igualdad, seguridad y control. Incluso los parámetros aparentemente técnicos pueden contener ya una toma de postura sobre qué tipo de daño colateral resulta aceptable en el ejercicio del poder punitivo.

Comprendida en su conjunto, la herencia del algoritmo permite afirmar que los sistemas de inteligencia artificial implementados en la política criminal no son dispositivos vacíos, sino artefactos cargados de memoria institucional. Heredan datos producidos a partir de prácticas desiguales de observación, categorías que organizan la realidad desde la lógica del riesgo, prioridades político-criminales que orientan la intervención, trayectorias históricas de vigilancia diferencial y márgenes normativos de tolerancia al error. Esa acumulación de herencias explica por qué el sesgo algorítmico no debe entenderse como una anomalía contingente, sino como una consecuencia posible de la estructura que configura al sistema desde su origen.

Precisamente por ello, la herencia del algoritmo no es accesoria del análisis, debe referirse al punto desde el cual comienza a hacerse inteligible la producción de sesgo en

sentido estricto. El sesgo no aparece simplemente al final, cuando el sistema produce un resultado injusto o discriminatorio. Comienza a incubarse mucho antes, en la forma en que se construyen los datos, se definen las categorías, se fijan las prioridades y se aceptan determinados costos de error. Si esto es así, entonces el sesgo algorítmico no debe pensarse como una anomalía contingente que ocasionalmente corrompe herramientas en principio neutrales, sino como una consecuencia posible de la herencia estructural que las configura desde su origen. Ése será, precisamente, el siguiente problema a desarrollar.

V. Efectos de la herencia algorítmica: sesgo, selectividad y racionalidad punitiva

Una vez identificadas las capas de herencia que configuran al algoritmo, corresponde examinar la manera en que éstas se proyectan sobre el funcionamiento efectivo del control penal. En este punto, el sesgo algorítmico importa no sólo por la incorrección de ciertos resultados, sino porque incide en la distribución de vigilancia, sospecha e intervención institucional. Por tanto, lo relevante no es únicamente que el sistema clasifique, sino cómo esa clasificación reorganiza de manera desigual la exposición al poder punitivo.

Conviene distinguir, en este sentido, entre al menos tres planos del sesgo: el de los datos, cuando la información de entrada refleja historias desiguales de observación e intervención; el del diseño, cuando las variables y categorías incorporan presupuestos problemáticos sobre riesgo o sospecha; y el de la implementación, cuando el uso institucional del sistema profundiza asimetrías ya existentes. La importancia de esta distinción radica en mostrar que el sesgo puede incubarse en distintos momentos del ciclo algorítmico y no sólo en el resultado final de la herramienta.

En el ámbito penal, esta cuestión reviste especial gravedad porque el sesgo no se traduce únicamente en una mala decisión técnica, sino en una distribución desigual de sospecha, vigilancia y exposición al aparato coercitivo del Estado. Un algoritmo que prioriza ciertos espacios identifica determinados perfiles como más riesgosos o asigna mayores probabilidades de reincidencia a poblaciones previamente sobrerrepresentadas, no sólo produce una clasificación discutible, sino que reordena el mapa práctico del control penal. Allí donde el sistema señala riesgo o anomalía, tiende a intensificarse la intervención institucional; donde no lo hace, la atención puede relajarse.

Esta constatación permite advertir con mayor nitidez la relación entre sesgo algorítmico y selectividad penal. La selectividad funciona como uno de sus rasgos constitutivos, ya que el poder punitivo jamás se distribuye de manera homogénea sobre el conjunto social, sino que históricamente se concentra sobre ciertos sectores, espacios y subjetividades (Baratta, 1986). Lo novedoso en el escenario algorítmico resulta en la posibilidad de automatizarla, racionalizarla y presentarla desde una apariencia de objetividad técnica. El riesgo consiste precisamente en que la mediación tecnológica vuelva menos visibles las decisiones de exclusión o focalización, al recubrirlas con el lenguaje de la predicción, la optimización y la eficiencia. Así, la selectividad deja de percibirse como una operación política discutible y comienza a presentarse como una consecuencia casi natural de lo que "muestran los datos".

Por esa razón, uno de los efectos más delicados de los sistemas algorítmicos es la sobrerrepresentación de ciertos grupos y territorios. Si determinados espacios han sido históricamente objeto de una observación institucional más intensa, es probable que generen también una mayor

densidad de datos de contacto con el sistema. Esa acumulación puede ser leída luego por el algoritmo como evidencia de mayor riesgo, lo que justificaría nuevas intervenciones y cerraría un circuito autorreforzado de vigilancia y clasificación.

Esto obliga a tomar en serio la dimensión indirecta de la discriminación algorítmica. En muchos casos, el sistema no requiere incorporar variables abiertamente prohibidas para producir resultados desiguales; basta con que opere sobre indicadores funcionalmente correlacionados con condiciones estructurales de desventaja. De ahí que la crítica, aparte de verificar si existe una categoría explícitamente discriminatoria, deberá examinar el conjunto de relaciones que permite que ciertas diferencias socialmente sensibles reingresen al sistema con una apariencia técnicamente neutral.

Sin embargo, reducir el problema a la producción de sesgos o a la automatización de la selectividad sería insuficiente. Tales efectos se insertan en una racionalidad punitiva más amplia, orientada a anticipar, clasificar y administrar de manera diferencial la intervención penal. En ese marco, el algoritmo contribuye a consolidar una determinada forma de gobierno del riesgo.

De acuerdo con lo anterior, la clasificación de sujetos y poblaciones adquiere una centralidad decisiva. En el terreno penal, clasificar nunca ha sido una simple operación descriptiva: implica distribuir grados de vigilancia, escrutinio y tolerancia institucional. Cuando esa facultad se amplifica mediante modelos algorítmicos, el sistema puede volverse más eficiente para focalizar intervenciones, pero también más proclive a consolidar mecanismos de sospecha permanente sobre sujetos previamente expuestos al control.

La racionalidad que emerge de estos procesos no necesita apelar de modo abierto a categorías clásicas de peligrosidad para producir efectos equivalentes. Le basta con modular las intensidades de intervención y desplazar hacia ciertos sectores sociales los costos de una vigilancia más persistente y de una menor presunción de normalidad. De este modo, la inteligencia artificial consolida nuevas herramientas al sistema penal, mientras que refuerza aquella política criminal que vaticinaba Eugenio Raúl Zaffaroni cada vez más cómoda con la clasificación preventiva y la desigual distribución de la exposición al control como una expresión de ejercicio selectivo del poder punitivo (Zaffaroni, 1998).

Visto así, el principal problema no es que el algoritmo se equivoque, sino que sus resultados puedan consolidarse dentro de estructuras institucionales ya marcadas por la selección desigual y la anticipación preventiva. Un sistema preciso, pero construido sobre bases sesgadas o inserto en contextos históricamente desiguales no corrige necesariamente la injusticia del control penal, sino que puede hacerla más estable y menos visible. Por ello, el examen de estos efectos no puede detenerse en la descripción del sesgo, la sobrerrepresentación o la racionalidad punitiva que los sostiene. Es necesario dar un paso adicional y preguntarse, si es que ello resulta posible, en qué condiciones el uso de IA en la política criminal puede someterse a límites compatibles con un Estado democrático y con un modelo garantista del poder punitivo.

VI. Límites democráticos y garantistas del uso de la IA en la política criminal

Si el análisis precedente ha mostrado que la inteligencia artificial puede reforzar sesgos heredados, automatizar formas de selectividad y robustecer una racionalidad punitiva orientada a la anticipación, el problema final consiste en establecer cuáles son los límites normativos e institucionales que deben imponerse para evitar que la promesa de eficiencia legitime nuevas formas de opacidad, desigualdad y expansión del control penal. Nuestra conclusión es alejarnos del pensamiento errático de prescindir de la tecnología y precisar en qué condiciones su empleo podría resultar compatible con un Estado democrático de derecho. En materia penal, esta exigencia es especialmente intensa, porque la incorporación de sistemas algorítmicos incide en el espacio más sensible de intervención estatal sobre la libertad, la igualdad, la privacidad, la presunción de inocencia y el debido proceso. Precisamente por ello, su validez no puede medirse únicamente por su capacidad predictiva, su rendimiento operativo o su utilidad administrativa.

Una primera exigencia democrática consiste en desplazar el entusiasmo tecnocrático que suele acompañar a estas herramientas. En el ámbito penal, la eficacia no constituye por sí sola un criterio suficiente de legitimidad. Un sistema puede ser eficiente para focalizar patrullajes, identificar patrones, priorizar expedientes o segmentar perfiles de riesgo y, sin embargo, resultar incompatible con los principios que limitan el ejercicio del poder punitivo. Esta distinción es fundamental porque obliga a recordar que el problema de la IA no se agota en lo que puede hacer, sino que depende de lo que jurídicamente debe permitírsele hacer. En una política criminal democrática, la incorporación de nuevas capacidades tecnológicas no puede entenderse

como una autorización automática para ampliar la vigilancia, flexibilizar estándares de intervención o debilitar controles institucionales. Por el contrario, cuanto mayor es la capacidad de una herramienta para producir clasificaciones, inferencias y afectaciones masivas, mayores deben ser también las garantías que regulen su diseño, supervisión y uso.

En este punto, la transparencia ocupa un lugar central, aunque debe ser entendida en un sentido más exigente que la mera publicidad formal de la existencia del sistema. No basta con que una institución reconozca que utiliza inteligencia artificial (Unión Europea, 2024), es indispensable que puedan conocerse, al menos en un grado suficiente para el control democrático, las finalidades concretas de la herramienta, el tipo de decisiones para las que se emplea, las fuentes de datos que la alimentan, las categorías relevantes de clasificación, los márgenes de error que se consideran aceptables y las instancias responsables de su validación y supervisión. La opacidad algorítmica constituye un problema técnico y una dificultad político-jurídica de primer orden porque allí donde no puede saberse cómo opera una herramienta que influye en decisiones de vigilancia, investigación o priorización penal, también se debilita la posibilidad de impugnar sus efectos, exigir rendición de cuentas y reparar eventuales afectaciones a derechos fundamentales.

Sin embargo, la transparencia por sí sola tampoco resuelve el problema. En el ámbito penal se requiere, además, trazabilidad. Es decir, la posibilidad de reconstruir el recorrido de la decisión algorítmica y de identificar cómo intervienen en ella los datos de entrada, las reglas de procesamiento, los criterios de clasificación y las decisiones humanas asociadas a su implementación. La trazabilidad

cumple así una función estrictamente garantista: restituye el vínculo entre la decisión y la responsabilidad pública, impide que la autoridad se oculte detrás del lenguaje técnico y permite someter el uso de la herramienta a escrutinio jurídico y político.

A esto debe añadirse una exigencia más profunda: la evaluación de impacto en derechos humanos. Si los sistemas de inteligencia artificial operan sobre contextos históricamente marcados por selectividad penal, vigilancia desigual y discriminación estructural, su adopción no puede justificarse sin una valoración previa y posterior de sus posibles efectos sobre la igualdad, la no discriminación, la privacidad, la libertad personal, la tutela judicial efectiva y el debido proceso. En otras palabras, las instituciones deberán revisar si sus resultados son compatibles con los compromisos constitucionales e internacionales en materia de derechos humanos. Esta evaluación no puede ser un trámite decorativo ni un discurso *ex post* de legitimación. Debe implicar un análisis riguroso de los riesgos previsibles del sistema, de las poblaciones que podrían resultar especialmente afectadas, de los sesgos que podrían reproducirse y de los mecanismos correctivos o prohibitivos que corresponde activar antes de su despliegue operativo.

La supervisión pública e independiente constituye otro límite indispensable. En materia penal no basta con confiar en que las propias instituciones usuarias del sistema realicen un autocontrol suficiente. La experiencia comparada muestra que las herramientas tecnológicas, cuando se insertan en aparatos de seguridad, vigilancia y persecución, tienden a quedar protegidas por narrativas de necesidad, reserva o especialización técnica que dificultan el escrutinio externo, tal como sucedió después del 11S (Lyon, 2007). De ahí que una política criminal democrática deba prever mecanismos

de auditoría independientes, controles judiciales cuando corresponda, acceso de órganos públicos de supervisión y, en ciertos supuestos, posibilidades de revisión por parte de la academia, la sociedad civil especializada o las entidades autónomas con capacidades técnicas reales. La IA no puede convertirse en una zona exenta de control por el simple hecho de estar asociada a procesos complejos o a lenguajes matemáticos de difícil acceso. Mientras más intensa sea su incidencia en decisiones de vigilancia o intervención penal, más robustos deben ser los controles institucionales sobre su legitimidad y funcionamiento.

Inevitablemente, esta discusión conduce a una cuestión decisiva: no todos los usos de la inteligencia artificial en el ámbito penal son equivalentes y, por ello, tampoco todos son igualmente admisibles. Existe una diferencia relevante entre herramientas orientadas a tareas administrativas o de apoyo no decisivo y aquellas que inciden directamente en decisiones sensibles sobre vigilancia, clasificación de sujetos, investigación, valoración del riesgo, priorización coercitiva o afectación de derechos. Cuanto más cercano se encuentre el sistema al núcleo duro del poder punitivo, más severa debe ser la exigencia de justificación y más estrictos los límites para su empleo. En algunos casos, el problema no será simplemente de regulación, sino de incompatibilidad material con un modelo garantista. Ello puede ocurrir, por ejemplo, cuando la herramienta produce clasificaciones opacas imposibles de controvertir, cuando reposa sobre bases de datos marcadamente discriminatorias, cuando amplifica de manera desproporcionada la vigilancia sobre ciertos sectores o cuando su funcionamiento erosiona de manera estructural la presunción de inocencia, el derecho de defensa y, por lo tanto, el debido proceso.

Desde una perspectiva garantista, esto significa que el uso de inteligencia artificial en política criminal debe someterse a una lógica de restricción y no de expansión automática. La pregunta correcta no es cómo aprovechar al máximo las capacidades disponibles, sino qué límites deben imponerse para impedir que la tecnología erosione principios que históricamente han surgido precisamente para contener los abusos del castigo. El garantismo penal recuerda, en este sentido, que el poder punitivo sólo es legítimo cuando se encuentra normativamente encauzado, sometido a controles estrictos y limitado por derechos fundamentales indisponibles. Aplicado al campo algorítmico, ello obliga a rechazar cualquier idea, según la cual la mera capacidad técnica de anticipar, clasificar o correlacionar baste para justificar la intervención estatal. La posibilidad de hacerlo no equivale a la licitud de hacerlo, ni la utilidad institucional reemplaza la necesidad de fundamentación jurídica. En esa línea, los criterios de equidad algorítmica pueden ofrecer parámetros mínimos de contención. Uno de ellos es la paridad estadística que, para Kearns y Roth supone que la proporción de resultados favorables otorgados por un sistema se mantenga aproximadamente igual entre grupos protegidos, de modo que la pertenencia a uno u otro no altere significativamente la tasa de aprobación o beneficio (Kearns, 2020). Trasladada al campo de la política criminal, esta exigencia obliga a interrogar si la tasa de decisiones positivas o adversas varía sustancialmente entre grupos definidos por atributos sensibles, como raza o sexo, y si tales diferencias responden a criterios jurídicamente admisibles o a la reproducción encubierta de desigualdades estructurales que se identifican como grupos definidos por atributos sensibles, como por ejemplo raza o sexo.

A partir de esta lógica, una política criminal democrática frente a sistemas algorítmicos exige al menos cuatro cautelas de fondo. La primera consiste en reconocer que la inteligencia artificial no corrige por sí misma las desigualdades del sistema penal y que, en contextos de herencias institucionales sesgadas, puede incluso consolidarlas con formas más sofisticadas. La segunda exige no confundir objetividad computacional con neutralidad normativa, pues el algoritmo siempre llega mediado por decisiones humanas previas sobre datos, categorías, prioridades y fines de intervención. La tercera impone que toda herramienta tecnológicamente compleja permanezca sometida a responsabilidad pública, lo que excluye la delegación opaca de funciones estatales en modelos no escrutables. La cuarta obliga a preservar, frente a cualquier innovación, la centralidad de los derechos humanos como criterio rector de admisibilidad, control y eventual prohibición.

Desde esta perspectiva, el aporte principal de la noción de genética del algoritmo puede apreciarse con mayor claridad al cierre del artículo. Su utilidad permite una comprensión más profunda de por qué los sistemas de inteligencia artificial aplicados a la política criminal no deben ser evaluados sólo por sus resultados inmediatos. La genética del algoritmo muestra que tales sistemas incorporan memorias institucionales, lenguajes clasificatorios, prioridades de vigilancia y márgenes de tolerancia al error que condicionan desde el origen su funcionamiento y sus efectos. En consecuencia, la crítica no puede permanecer en la superficie de la precisión estadística o de la falla puntual, sino que debe dirigirse al entramado político, normativo e histórico que alimenta el dispositivo y le otorga capacidad de intervenir sobre sujetos y poblaciones concretas.

A partir de ello, es posible sostener como conclusión general que la IA además de ser una innovación instrumental en el ámbito penal, también representa un punto de intensificación de problemas previos de selectividad, opacidad y expansión del control. Su novedad técnica no elimina la continuidad con viejas formas de clasificación desigual, vigilancia concentrada y administración diferencial del riesgo; por el contrario, puede volverlas más eficaces, más estables y menos visibles, como mencionamos anteriormente. Precisamente ahí reside la relevancia del concepto propuesto: en mostrar que el algoritmo procesa información heredada de una forma históricamente construida de mirar, ordenar y administrar el desorden social. Esa herencia importa porque condiciona quién será visto, cómo será clasificado, con qué intensidad será observado y bajo qué umbral de sospecha quedará expuesto al aparato penal.

Por ello, una política criminal compatible con el Estado constitucional no puede incorporar sistemas de inteligencia artificial desde la fascinación técnica ni desde la lógica de la mera optimización institucional. Debe hacerlo, en todo caso, desde una cautela reforzada, consciente de que estas herramientas operan sobre un campo atravesado por desigualdades estructurales y por una historia larga de selectividad punitiva. La pregunta decisiva es si la IA y su uso en la política criminal puede someterse a límites suficientemente robustos como para impedir que reproduzca o profundice injusticias ya arraigadas en las instituciones. Allí se juega, en definitiva, el verdadero alcance democrático de la discusión.

En suma, la genética del algoritmo constituye una categoría útil para pensar críticamente la relación entre inteligencia artificial y política criminal porque permite articular tres planos que con frecuencia aparecen disocia-

dos: la procedencia estructural de los sistemas, sus efectos prácticos sobre la distribución desigual del control penal y la necesidad de imponer límites garantistas a su utilización. Entendida así, la crítica de la inteligencia artificial en el ámbito penal debe someterla a un examen riguroso de procedencia, legitimidad y compatibilidad con los derechos fundamentales. Sólo con ese escrutinio será posible evitar que, detrás del lenguaje de la innovación, se consoliden nuevas formas de racionalidad punitiva menos visibles, pero no por ello menos intensas.

VII. Referencias

Arteaga, Nelson y Javier Arzuaga (2017). *Sociologías de la violencia: estructuras, sujetos, interacciones y acción simbólica*. México: Flacso.

Baratta, Alessandro (1986). *Criminología crítica y crítica del derecho penal: Introducción a la sociología jurídico-penal*. México: Siglo XXI.

Chabert, Jean Luc (ed.) (1994). *A history of algorithms. From the Pebble to the Microchip*. París: Belin.

Clark, Ronald (1997). *Situational Crime Prevention: Succesful Case Studies*. New York: Harrow and Heston.

Crawford, Kate (2021). *Atlas para IA*. Buenos Aires: Yale University Press.

Feeley, Malcolm and Jonathan Simon (1992). *The new penology. Notes on the Emerging Strategy of Corrections and Its Implications. Criminology*, 30(4), 449-474.

Ferrajoli, Luigi (1995). *Derecho y razón. Teoría del Garantismo Penal*. Madrid: Trotta.

Kearns, Michael y Aaron Roth (2020). *El algoritmo ético*. Madrid: Woltors Kluwer España.

Lyon, David (2007). *Surveillance Studies: An Overview*. Cambridge: Polity Press.

O´Neil, Cathy (2016). *Wapons of math estruction*. New York: Broadway books.

UE: Unión Europea (2024, 13 de junio). Regulation of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (Artificial Intelligence Act). *Official Journal of the European Union*, 1689. Publications Office of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Zaffaroni, Eugenio (1998). *En busca de las penas perdidas. Deslegitimación y dogmática jurídico penal*. Buenos Aires: Ediar.