EDUCA - PLUS
EDITORIAL ACADÉMICA

CRIMINALÍSTICA DEL FUTURO



Octubre 2026 – Educa Plus Editorial Académica
Copyright © Educa Plus Editorial Académica
Copyright del texto © 2026 de Autores
Ecuador, Naranjito, Cod. Post. 091803
Email: editorial@educapluseditorial.com
Contacto: +593 960 838 709
https://biblioteca.educapluseditorial.com/

Datos Técnicos de Publicación Internacional
Título: Criminalística del futuro
Subtítulo: Ciencia, inteligencia artificial y transformación forense
Autores: Erazo Parra, J. L., de la Torre Núñez, E. del C., Sánchez Pólit, F. R., Cáceres Manzano, V. P., Lucena de Ustáriz, M. E., & Ustáriz Fajardo, F. J.
Diseño y producción editorial: Durán Sánchez, Stephany
Editor: Barco Acosta, Jonathan
Formato: PDF
Páginas: 146
Tamaño: A4 21x29.7cm
Modo de acceso: Recurso digital disponible en línea
ISBN: 978-9907-9622-6-0
DOI: https://doi.org/10.66646/educa-plus/10oct2627
Prefijo Editorial ISBN: 978-9907-9571

Este libro fue sometido a arbitraje académico bajo el sistema de evaluación por pares doble ciego (*double-blind peer review*).

Disponible para su descarga gratuita en https://biblioteca.educapluseditorial.com/

EDITORIAL

Disponible para su descarga gratuita en https://biblioteca.redsa.ec/

Los contenidos de este libro pueden ser descargados, reproducidos, difundidos e impresos con fines de estudio, investigación y docencia, o para su utilización en productos o servicios no comerciales, siempre que se reconozca adecuadamente a los autores como fuente y titulares de los derechos de propiedad intelectual, sin que ello implique en modo alguno que aprueban las opiniones, productos o servicios resultantes. En el caso de contenidos que indiquen expresamente que proceden de terceros, deberán dirigirse a la fuente original indicada para gestionar los permisos correspondientes.

Cómo citar esta obra

Erazo Parra, J. L., de la Torre Núñez, E. del C., Sánchez Pólit, F. R., Cáceres Manzano, V. P., Lucena de Ustáriz, M. E., & Ustáriz Fajardo, F. J. (2026). Criminalística del futuro: Ciencia, inteligencia artificial y transformación forense. Educa Plus Editorial Académica. https://doi.org/10.66646/educa-plus/10oct2627

Proceso Editorial:

Cada uno de los textos publicados por Educa Plus Editorial Académica ha sido sometido a un riguroso proceso de evaluación por pares doble ciego externos (double-blind peer review), conforme a los estándares editoriales internacionales y a la normativa interna de la editorial.

Aviso Legal:

La información presentada, así como el contenido, fotografías, gráficos, cuadros, tablas y referencias de este manuscrito, es de exclusiva responsabilidad del/los autor/es y no refleja necesariamente el pensamiento ni la posición institucional de la Educa Plus Editorial Académica. Las opiniones expresadas en cada capítulo son responsabilidad exclusiva de sus respectivos autores.

Derechos de Autor ©

Este documento se publica bajo los términos y condiciones de la licencia:

Creative Commons Reconocimiento — No Comercial — Compartir Igual 4.0 Internacional (CC BY-NC-SA 4.0)

Usted es libre de compartir y adaptar el material, siempre que otorgue el crédito correspondiente, no lo utilice con fines comerciales y distribuya sus contribuciones bajo la misma licencia.

Para ver una copia de esta licencia, visite: https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es

AGRADECIMIENTOS

A quienes, con paciencia y rigor, siguen creyendo que la verdad puede reconstruirse a partir de un indicio.

A los peritos, investigadores y operadores de justicia de Iberoamérica, cuyo trabajo cotidiano, muchas veces silencioso, sostiene la posibilidad de una justicia basada en evidencia.

CONTENIDO

Agradecimientos 5
Presentación 10

CAPÍTULO 1 12

De Locard al dato: la evolución de la prueba científica

1.1. Caso de apertura: cuando la ciencia forense falla 13
1.2. Introducción: qué entendemos por criminalística 13
1.3. Los cimientos de la disciplina: de Bertillon a Locard 14
1.4. La consolidación del siglo veinte: entre el rigor y la costumbre 15
1.5. La crisis de validez científica: los informes NAS y PCAST 16
1.6. El factor humano: sesgo cognitivo en el peritaje forense 17
1.7. Respuestas institucionales: hacia una criminalística basada en evidencia 19
1.8. La criminalística en Iberoamérica: entre la reforma procesal y el rezago tecnológico 20
1.9. El umbral de la inteligencia artificial: por qué esta transformación es distinta 21
1.10. Riesgos y buenas prácticas 22

CAPÍTULO 2 23

Inteligencia Artificial para el forense: fundamentos

2.1. Caso de apertura: cuando el algoritmo también se equivoca 24
2.2. Qué entendemos por inteligencia artificial en el contexto forense 24
2.3. Aprendizaje automático: conceptos fundamentales 26
2.4. Aprendizaje profundo y redes neuronales 27
2.5. Visión computacional aplicada a la evidencia forense 28
2.6. Procesamiento del lenguaje natural y lingüística forense computacional 28
2.7. Inteligencia artificial generativa: creación y detección de contenido sintético 30
2.8. El problema de la caja negra: explicabilidad e interpretabilidad 31
2.9. Qué puede y qué no puede hacer hoy un algoritmo con evidencia forense 33
2.10. Riesgos y buenas prácticas 33

CAPÍTULO 3 35

Identificación humana aumentada: biometría, reconocimiento facial, AFIS con IA, antropología y odontología forense

3.1. Caso de apertura: el caso Mayfield y los límites de la certeza dactiloscópica 36
3.2. Qué entendemos por identificación humana aumentada 37
3.3. Biometría dactilar: de AFIS al aprendizaje profundo 38
3.4. Reconocimiento facial forense: verificación, identificación y sus límites 41
3.5. Antropología forense asistida: estimación de sexo y edad mediante inteligencia artificial 42
3.6. Odontología forense e inteligencia artificial: entre la estimación de edad y la cautela sobre las marcas de mordida 44
3.7. Fusión multimodal y bases de datos biométricas interoperables 45

3.8. Riesgos y buenas prácticas 46

CAPÍTULO 4 48

Genética forense de nueva generación: secuenciación masiva, fenotipado forense por ADN y genealogía genética investigativa

4.1. Caso de apertura: el caso del Golden State Killer y el nacimiento de la genealogía genética investigativa 49
4.2. La genética forense más allá del perfil STR 50
4.3. Secuenciación masiva paralela: qué cambia respecto a la genética forense clásica 50
4.4. Genotipado probabilístico y mezclas complejas de ADN 52
4.5. Fenotipado forense por ADN: prediciendo la apariencia física a partir de una muestra 53
4.6. Genealogía genética investigativa: bases de datos, métodos y controversias 55
4.7. Ética, privacidad y gobernanza de la genética forense de nueva generación 57
4.8. Riesgos y buenas prácticas 58

CAPÍTULO 5 60

La escena del crimen digital: escaneo láser 3D, fotogrametría, drones y reconstrucción virtual

5.1. Caso de apertura: cuando lo más avanzado no es necesariamente lo más convincente 61
5.2. La escena del crimen como objeto digital 62
5.3. Escaneo láser tridimensional: fundamentos y aplicaciones 63
5.4. Fotogrametría: reconstrucción tridimensional a partir de imágenes 65
5.5. Drones y vehículos aéreos no tripulados en la investigación forense 66
5.6. Reconstrucción virtual y presentación ante tribunales 68
5.7. Desafíos: costo, estandarización, formación y admisibilidad 69
5.8. Riesgos y buenas prácticas 70

CAPÍTULO 6 72

Rastros materiales y algoritmos: balística, documentoscopia y análisis de trazas asistidos por IA

6.1. Caso de apertura: la balística forense responde a la crisis de validez 73
6.2. La naturaleza de los rastros materiales: entre la comparación visual y el puntaje cuantificado 74
6.3. Balística forense e inteligencia artificial: comparación automatizada de casquillos y proyectiles 75
6.4. El problema de las decisiones inconclusas y la interpretación algorítmica de puntajes 76
6.5. Documentoscopia: análisis forense de escritura y firmas asistido por inteligencia artificial 77
6.6. Análisis de trazas: fibras, vidrio, pintura y residuos de incendio 79
6.7. Un hilo común: de la subjetividad al puntaje cuantificado 80
6.8. Riesgos y buenas prácticas 81

CAPÍTULO 7 **83**

El nuevo frente digital: informática forense ante deepfakes, criptoactivos, IoT y los límites del análisis predictivo

7.1. Caso de apertura: cuando los detectores de deepfakes se enfrentan al mundo real 84
7.2. El nuevo perímetro de la evidencia digital 84
7.3. Deepfakes y contenido sintético: detección forense en la práctica real 86
7.4. Criptoactivos: el mito de la anonimidad y la forense de blockchain 87
7.5. Internet de las Cosas: la escena del crimen distribuida 89
7.6. Los límites reales de la inteligencia predictiva y el análisis criminal 90
7.7. Un hilo común: la velocidad del cambio tecnológico frente a la velocidad de la validación 92
7.8. Riesgos y buenas prácticas 93

CAPÍTULO 8 **95**

Validez y derecho: la prueba algorítmica ante el tribunal

8.1. Caso de apertura: el primer video mejorado por inteligencia artificial ante un tribunal penal 96
8.2. Qué significa prueba algorítmica ante un tribunal 97
8.3. Estándares de validación: la serie ISO 21043 y la acreditación de herramientas de inteligencia artificial 98
8.4. El estándar Daubert y sus factores aplicados a sistemas de inteligencia artificial 99
8.5. Explicabilidad como requisito de admisibilidad 101
8.6. Cadena de custodia digital en la era algorítmica 102
8.7. La brecha de recursos: acceso a expertos y defensa efectiva 103
8.8. Sistemas acusatorios latinoamericanos: entre la aceptación general y el escrutinio Daubert 104
8.9. Riesgos y buenas prácticas 105

CAPÍTULO 9 **107**

Ética, sesgo y derechos humanos: sesgo algorítmico, privacidad y marcos regulatorios

9.1. Caso de apertura: Machine Bias y el nacimiento del debate sobre sesgo algorítmico en la justicia penal 108
9.2. De la validez técnica a la justicia algorítmica 109
9.3. Sesgo algorítmico: tipos, fuentes y el dilema de las métricas de equidad en competencia 110
9.4. Privacidad y vigilancia biométrica: el marco europeo como referencia regulatoria 112
9.5. Marcos regulatorios en América Latina: entre la aspiración y la implementación 114
9.6. Hacia un marco ético transversal para la criminalística con inteligencia artificial 116
9.7. Riesgos y buenas prácticas 117

CAPÍTULO 10 **119**

El laboratorio 4.0 y el perito del futuro: gestión del cambio institucional, nuevas competencias y prospectiva 2035-2050

10.1. Caso de apertura: lo que reveló el escándalo de Colorado sobre la gestión institucional del cambio 120
10.2. El laboratorio 4.0 como transformación organizacional, no solo tecnológica 121
10.3. Gestión del cambio institucional: lecciones de la transformación digital en laboratorios forenses 122
10.4. La crisis de capital humano: escasez, acumulación de casos pendientes y el papel de la inteligencia artificial 123
10.5. El perito del futuro: nuevas competencias y alfabetización algorítmica 124
10.6. Prospectiva 2035-2050: escenarios posibles para la criminalística 126
10.7. Cooperación regional: una vía para acelerar el escenario de integración madura 128
10.8. Riesgos y buenas prácticas 129

Epílogo 131
Bibliografía 133

PRESENTACIÓN

Este libro nació de una pregunta que cualquier perito, fiscal, juez o defensor de Iberoamérica se formula hoy, con distinto grado de urgencia: ¿qué cambia, realmente, cuando la inteligencia artificial entra a un laboratorio forense? La respuesta que proponen las páginas siguientes es, deliberadamente, menos espectacular que la que suele ofrecer el debate público sobre esta tecnología. La inteligencia artificial no resuelve, por sí sola, los problemas históricos de la criminalística, ni los agrava de manera irremediable. Los reproduce, a una velocidad y una escala considerablemente mayores, y exige de la disciplina exactamente el mismo rigor que ya le exigieron, hace más de una década, los informes que documentaron su crisis de validez científica.

Ese es el hilo que atraviesa los diez capítulos de esta obra. El primero reconstruye la historia de la criminalística desde Edmond Locard hasta la crisis de validez de 2009 y 2016, y explica por qué esa historia es indispensable para entender lo que sigue. El segundo ofrece los fundamentos conceptuales de la inteligencia artificial que un profesional forense necesita dominar, sin pretender convertirlo en informático. Los capítulos tercero a séptimo examinan, disciplina por disciplina, cómo esta tecnología transforma la identificación humana, la genética forense, la escena del crimen, los rastros materiales y el nuevo frente digital de los deepfakes, los criptoactivos y el análisis predictivo. Los capítulos octavo y noveno abordan las dos exigencias que ninguna innovación tecnológica puede eludir en un proceso penal: la admisibilidad jurídica de la prueba y su compatibilidad con los derechos humanos. El décimo capítulo cierra el recorrido con una mirada institucional, dedicada a la gestión del cambio, a la formación del perito del futuro y a la prospectiva de la disciplina hacia 2035 y 2050.

Este libro está escrito pensando, de manera explícita, en el contexto iberoamericano. No porque los desafíos descritos en sus páginas sean exclusivos de la región, sino porque la literatura disponible sobre inteligencia artificial forense suele producirse y discutirse casi por completo desde fuera de ella. Cada capítulo dedica, por eso, un espacio explícito a la situación específica de los sistemas de justicia latinoamericanos: sus sistemas procesales acusatorios, sus institutos forenses, sus marcos regulatorios todavía incipientes y las poblaciones que sus laboratorios atienden y que rara vez están representadas en los datos con los que se entrenan las herramientas comerciales que la región termina adoptando.

El método de este libro es, en cada capítulo, el mismo: partir de un caso real y documentado, examinar la evidencia científica más reciente y rigurosa disponible sobre cada tecnología, y extraer de ese examen un conjunto de riesgos y buenas prácticas concretas. Cada capítulo cierra con preguntas de discusión pensadas para el aula, el seminario o la sala de reuniones de un laboratorio forense, y con una bibliografía verificable que privilegia, salvo excepciones clásicas plenamente justificadas, la investigación académica de los últimos cinco años.

Este libro no está dirigido únicamente a especialistas en inteligencia artificial, ni únicamente a criminalistas tradicionales, sino a ambos, y sobre todo al espacio que hoy los separa. Está pensado para peritos que necesitan entender qué pueden y qué no pueden exigirle a un algoritmo, para fiscales y defensores que deben valorar una prueba algorítmica sin ser ingenieros, para jueces que deben decidir sobre su admisibilidad, para estudiantes de criminalística que se forman en un momento de transformación acelerada de su propia disciplina, y para cualquier lector interesado en comprender, con rigor y sin sensacionalismo, qué significa buscar la verdad material de un hecho en la era de la inteligencia artificial.

Quede, entonces, planteada la invitación con la que se abre el primer capítulo de esta obra: recorrer,

junto con Locard y con quienes hoy continúan su trabajo desde un laboratorio, el camino que va del dato al algoritmo, sin perder nunca de vista que ese camino sigue teniendo, al final, la misma meta de siempre.



CAPÍTULO 1

De Locard al dato: la evolución de la prueba científica

1.1. Caso de apertura: cuando la ciencia forense falla

Durante décadas, el examen microscópico de cabellos fue presentado en los tribunales estadounidenses como una técnica capaz de asociar, con un alto grado de certeza, un cabello encontrado en la escena del crimen con un sospechoso concreto. Peritos del FBI testificaron en cientos de causas penales utilizando ese método. La revisión sistemática de esos casos, impulsada tras la publicación del informe del National Research Council de 2009, arrojó un resultado que sacudió a la comunidad forense: en más del noventa por ciento de los primeros doscientos cincuenta y siete casos analizados se identificó al menos un tipo de error en el testimonio pericial (Innocence Project, 2019). Varias de esas causas terminaron en condenas a personas que después fueron exoneradas mediante pruebas de ADN.

Este episodio no es un caso aislado ni una curiosidad histórica. Es el síntoma de un problema estructural que atraviesa buena parte de la historia de la criminalística: la distancia entre lo que un método parece capaz de demostrar y lo que la evidencia científica disponible realmente respalda. El propósito de este capítulo es reconstruir esa historia, desde el nacimiento de la disciplina con Edmond Locard hasta la crisis de validez científica que estalló a comienzos del siglo veintiuno, para entender por qué la irrupción de la inteligencia artificial en el trabajo forense no puede analizarse como una simple actualización tecnológica, sino como una respuesta, entre varias posibles, a un problema epistemológico que la profesión arrastra desde hace más de un siglo.

Lo relevante del caso de los cabellos no es únicamente la magnitud del error, sino el mecanismo que lo hizo posible durante décadas. Los peritos que testificaron no actuaron de mala fe. Fueron formados dentro de una cultura institucional que daba por válido un método porque llevaba mucho tiempo utilizándose, porque contaba con aceptación general dentro del gremio y porque nunca había sido sometido a una revisión externa rigurosa. Esa combinación de tradición profesional, ausencia de estudios de error y confianza excesiva en la experiencia acumulada del examinador es, como se verá a lo largo de este capítulo, un patrón que se repite en múltiples disciplinas forenses y que constituye el núcleo del problema que los informes de 2009 y 2016 pusieron finalmente sobre la mesa.

1.2. Introducción: qué entendemos por criminalística

La criminalística puede definirse como el conjunto de disciplinas científicas y técnicas aplicadas a la identificación, el análisis y la interpretación de indicios materiales vinculados a un hecho presuntamente delictivo, con el fin de auxiliar a la administración de justicia (Saferstein, 2020). Se trata de un campo esencialmente interdisciplinario que integra aportes de la química, la física, la biología, la medicina, la estadística, la ingeniería y, cada vez con mayor intensidad, la ciencia de datos y la informática (Girard, 2021). Esta amplitud disciplinar es a la vez su mayor fortaleza y una de sus principales fuentes de vulnerabilidad, porque cada subdisciplina forense se ha desarrollado históricamente con niveles muy desiguales de rigor experimental, de estandarización y de validación empírica.

Conviene distinguir, siguiendo a Verdú (2023), entre prueba pericial y prueba científica. La prueba pericial es la opinión que un experto formula ante el sistema de justicia a partir de su conocimiento especializado, mientras que la prueba científica exige, además, que el método empleado haya sido sometido a validación, que su tasa de error sea conocida y que sus resultados sean reproducibles por otros examinadores independientes. Gran parte de la historia reciente de la criminalística puede leerse como el intento, todavía incompleto, de que la primera converja plenamente con la segunda.

Esta distinción no es meramente académica. De ella depende, en la práctica cotidiana de los

tribunales, el peso probatorio que se asigna a un dictamen pericial y, en última instancia, la libertad de las personas juzgadas. Un dictamen puede estar redactado con absoluta corrección formal, con un lenguaje técnico impecable y con la firma de un profesional debidamente acreditado, y aun así carecer de respaldo científico suficiente si el método subyacente nunca fue sometido a un estudio riguroso de exactitud. Reconocer esa posibilidad, incómoda para cualquier profesión, es el primer paso necesario para comprender por qué la disciplina se encuentra hoy en un proceso de transformación tan profundo, y por qué ese proceso no puede limitarse a la incorporación de nuevas tecnologías sin abordar, al mismo tiempo, sus fundamentos epistemológicos.

A lo largo de este capítulo se seguirá un hilo argumental compuesto por cuatro momentos. El primero reconstruye el nacimiento de la criminalística como disciplina científica autónoma, con especial atención a la figura de Locard. El segundo describe la expansión y, al mismo tiempo, la fragmentación metodológica que experimentó el campo durante el siglo veinte. El tercero analiza en profundidad la crisis de validez científica desencadenada por los informes de 2009 y 2016, así como el papel del sesgo cognitivo en la interpretación de la evidencia. El cuarto momento sitúa esa crisis en el contexto específico de Iberoamérica y traza el puente hacia la irrupción de la inteligencia artificial, tema que estructura el resto de este libro.

1.3. Los cimientos de la disciplina: de Bertillon a Locard

El impulso fundacional de la identificación criminal moderna se atribuye a Alphonse Bertillon, quien en la Francia de finales del siglo diecinueve desarrolló un sistema de identificación antropométrica basado en la medición sistemática de rasgos corporales considerados invariables en la edad adulta. El sistema bertillonage dominó la identificación policial durante varias décadas, hasta que la dactiloscopia demostró ser más precisa y menos costosa de implementar (Moreno y Silva, 2020). En paralelo, Hans Gross, jurista austríaco, sentó las bases de lo que llamó la criminalística como campo autónomo, al proponer que la investigación criminal debía apoyarse en el método científico y no exclusivamente en el testimonio o la confesión.

Es en ese linaje donde se inscribe Edmond Locard, discípulo del criminólogo Alexandre Lacassagne en Lyon, quien en 1910 convenció a la policía local de cederle dos habitaciones de un ático para instalar el que se considera el primer laboratorio policial de criminalística del mundo. Locard formuló el principio que hoy lleva su nombre y que constituye, todavía, el axioma central de toda la disciplina: todo contacto deja una huella. En su obra fundacional sobre la investigación criminal y los métodos científicos, Locard sostuvo que ninguna actividad humana, y en particular ninguna infracción a la ley, puede producirse sin dejar algún tipo de rastro material (Locard, 1920). Ese principio de intercambio, aparentemente simple, es el fundamento lógico sobre el cual se sostienen disciplinas tan diversas como la dactiloscopia, la balística, la genética forense o, más recientemente, la informática forense.

Locard no fue únicamente un teórico. Introdujo el método de identificación dactiloscópica de doce puntos, sistematizó el análisis documental y la pericia caligráfica, y llegó a publicar un tratado de siete volúmenes sobre criminalística que sirvió de referencia obligada en Europa durante buena parte del siglo veinte. En 1929 fundó la Academia Internacional de Criminalística, cuyo desarrollo se vio interrumpido por la Segunda Guerra Mundial. La huella de Locard en la disciplina va más allá de sus contribuciones técnicas puntuales, ya que instaló la idea de que el laboratorio, y no únicamente el testimonio, debía ocupar un lugar central en la investigación criminal.

Resulta útil detenerse en la naturaleza precisa del principio de intercambio, porque su interpretación contemporánea suele simplificarlo en exceso. Locard no afirmó que todo contacto deje siempre una huella detectable, ni que toda huella detectable sea automáticamente útil para la investigación. Su

planteamiento fue más matizado: sostuvo que el intercambio de materia es una consecuencia física inevitable del contacto entre dos superficies, independientemente de que la tecnología disponible en un momento histórico determinado sea capaz de detectar, recuperar e interpretar esa huella. Esta distinción entre la existencia física del indicio y la capacidad técnica para revelarlo resulta esencial para entender la historia posterior de la criminalística, porque buena parte del progreso de la disciplina a lo largo del siglo veinte puede describirse precisamente como la ampliación progresiva de esa capacidad de detección, desde el uso de la lupa y el microscopio óptico hasta la espectrometría de masas, la secuenciación genética y, en la actualidad, los sistemas de visión computacional capaces de identificar patrones invisibles al ojo humano.

Conviene señalar también que Locard trabajó en un contexto histórico marcado por la emergencia de la psicología del testimonio y por una creciente desconfianza hacia la prueba puramente testimonial, entonces la base casi exclusiva de la investigación criminal. Su insistencia en el indicio material como fuente de prueba más confiable que la memoria o la declaración humana anticipó, con casi un siglo de antelación, buena parte de las críticas que la psicología cognitiva formularía más tarde sobre la fiabilidad del testimonio de testigos presenciales. Esa misma tensión, entre la prueba material y el juicio humano que la interpreta, reaparecerá de forma casi idéntica cuando, un siglo después, la disciplina deba decidir cuánta autonomía otorgar a un sistema de inteligencia artificial frente al criterio del perito humano que supervisa su funcionamiento.

1.4. La consolidación del siglo veinte: entre el rigor y la costumbre

A lo largo del siglo veinte, la criminalística se expandió en múltiples direcciones. La dactiloscopia se consolidó como método de identificación con bases estadísticas relativamente sólidas, aunque su aplicación práctica siguió dependiendo, en última instancia, del juicio subjetivo del examinador. La balística forense desarrolló protocolos de comparación de características individuales en proyectiles y vainas, mientras que la documentoscopia refinó las técnicas de análisis de escritura, tintas y soportes. El desarrollo de la reacción en cadena de la polimerasa en la década de 1980 y su posterior aplicación al análisis de perfiles genéticos transformó radicalmente el campo, al introducir por primera vez un método forense capaz de expresar sus conclusiones en términos probabilísticos rigurosos y de estimar formalmente su tasa de error.

Sin embargo, ese progreso técnico no fue acompañado, en la mayoría de las disciplinas forenses distintas del ADN, por un desarrollo equivalente en materia de validación científica. Numerosas técnicas de comparación de patrones, entre ellas el análisis de marcas de mordida, el cotejo de huellas de calzado o la identificación de armas de fuego mediante características de herramienta, se instalaron en la práctica judicial y policial sobre la base de la aceptación profesional y la experiencia acumulada de los examinadores, más que sobre estudios experimentales que midieran de manera sistemática su exactitud, su repetibilidad y su reproducibilidad. Esta brecha entre práctica consolidada y evidencia empírica es precisamente lo que desencadenaría, a comienzos de este siglo, la crisis de validez que se describe en el apartado siguiente.

Un factor adicional que explica esta asimetría es el origen institucional de buena parte de la criminalística tradicional. A diferencia de la genética forense, que emergió directamente de la investigación académica en biología molecular y heredó de ella una cultura de publicación revisada por pares y de réplica experimental, disciplinas como la dactiloscopia o la balística se desarrollaron principalmente dentro de agencias policiales, orientadas ante todo a resolver casos concretos y no necesariamente a producir conocimiento generalizable sometido a escrutinio científico externo. Esta genealogía institucional distinta ayuda a explicar por qué, durante décadas, la validación de estos métodos se apoyó más en la experiencia acumulada y en la aceptación dentro del propio

gremio que en el tipo de evidencia empírica que exige el método científico convencional.

La odontología forense ofrece un ejemplo particularmente ilustrativo de esta tensión. Durante buena parte del siglo veinte se sostuvo que la dentición humana era suficientemente singular como para permitir la identificación individual a partir de marcas de mordida dejadas sobre la piel de una víctima. Investigaciones posteriores demostraron que la piel humana se deforma de manera impredecible al recibir una mordida, lo que introduce un margen de distorsión que ningún examinador puede controlar completamente, y que la premisa misma de unicidad de la dentición nunca fue demostrada con el rigor estadístico que sí existe, por ejemplo, para los perfiles genéticos (President's Council of Advisors on Science and Technology, 2016). El caso de las marcas de mordida se convertiría, años más tarde, en uno de los ejemplos más citados por los informes NAS y PCAST para ilustrar los riesgos de dar por válido un método forense sin someterlo previamente a una validación empírica rigurosa.

1.5. La crisis de validez científica: los informes NAS y PCAST

El punto de inflexión se produjo en febrero de 2009, cuando el National Research Council de la Academia Nacional de Ciencias de Estados Unidos publicó el informe Strengthening Forensic Science in the United States: A Path Forward, elaborado a solicitud del Congreso estadounidense tras años de preocupación por la falta de estándares científicos uniformes en el sistema forense (National Research Council, 2009). El comité, integrado por diecisiete especialistas que sostuvieron ocho reuniones y escucharon a setenta expositores a lo largo de casi dos años, llegó a una conclusión que resultó demoledora para buena parte de la profesión: con la excepción del análisis de ADN nuclear, ningún método forense había demostrado de manera rigurosa, mediante estudios científicos sólidos, su capacidad de asociar con un alto grado de certeza un elemento de evidencia con una fuente o un individuo específico (National Research Council, 2009).

El informe formuló trece recomendaciones que incluían la creación de una entidad federal independiente para supervisar la investigación forense, la exigencia de acreditación obligatoria de laboratorios y certificación de examinadores, el desarrollo de protocolos estandarizados y, de manera transversal, el impulso de un programa nacional de investigación destinado a establecer los límites y la tasa de error de cada disciplina forense (National Research Council, 2009). El impacto del informe trascendió las fronteras de Estados Unidos y se convirtió, según coinciden distintos analistas, en uno de los documentos más influyentes jamás publicados sobre la materia, al instalar de forma definitiva en el debate profesional conceptos como incertidumbre, probabilidad de coincidencia casual, fiabilidad y reproducibilidad (Forensic Statistics Working Group, 2019).

Siete años más tarde, en septiembre de 2016, el President's Council of Advisors on Science and Technology, conocido por sus siglas en inglés PCAST, publicó un segundo informe que profundizó y precisó los criterios de validación científica aplicables a lo que denominó métodos forenses de comparación de características (President's Council of Advisors on Science and Technology, 2016). El informe PCAST sostuvo que la validez fundacional de un método de este tipo solo puede establecerse mediante estudios independientes de caja negra, diseñados para medir con qué frecuencia los examinadores alcanzan conclusiones correctas al enfrentarse a un número suficiente de comparaciones representativas del uso real del método. Al aplicar ese criterio a seis disciplinas, entre ellas el análisis de mezclas complejas de ADN, las marcas de mordida, las huellas dactilares latentes, la identificación de armas de fuego y el análisis de calzado, PCAST concluyó que la mayoría de ellas carecía todavía de una base empírica suficiente, y que el análisis de marcas de mordida, en particular, no contaba con ningún estudio que respaldara su validez científica (President's Council of Advisors on Science and Technology, 2016) (véase la Figura 1.1).



Figura 1.1. Estado de la validez científica de los métodos forenses de comparación de características según el informe PCAST. Obsérvese que en 2016 solo el ADN de fuente única o de mezclas simples y las huellas latentes alcanzaron validez fundacional. Fuente: elaboración propia a partir de President's Council of Advisors on Science and Technology (2016).

El informe PCAST generó una controversia considerable dentro del propio sistema de justicia estadounidense. El Departamento de Justicia cuestionó públicamente que la validez científica de un método solo pudiera demostrarse mediante ese único tipo de diseño experimental, y sostuvo que la exigencia resultaba desproporcionada respecto de los estándares aplicados en otras áreas de la ciencia forense (Hunt, 2018). PCAST respondió en enero de 2017 con un addendum que revisó estudios adicionales aportados por organizaciones profesionales, sin modificar sustancialmente sus conclusiones centrales (President's Council of Advisors on Science and Technology, 2017). Más allá de cómo se resuelva esa disputa técnica, ambos informes coinciden en un diagnóstico compartido que sigue vigente: la criminalística necesita, con urgencia, un cuerpo de investigación empírica propio que permita conocer con precisión los límites y la tasa de error de cada uno de sus métodos.

1.6. El factor humano: sesgo cognitivo en el peritaje forense

Un segundo eje de la crisis de validez, distinto pero estrechamente vinculado al primero, es el papel que juega el sesgo cognitivo en la interpretación de la evidencia forense. El psicólogo cognitivo Itiel Dror, de la University College London, es hoy la referencia obligada en esta línea de investigación. Sus estudios, junto con los de Kassin, Kukucka y otros colegas, demostraron de manera consistente que la exposición a información contextual irrelevante, como el conocimiento de una confesión, la opinión de un investigador o antecedentes del sospechoso, induce a los examinadores forenses a emitir juicios sesgados en favor de la confirmación de una hipótesis previa (Kassin et al., 2013).

En un estudio particularmente influyente, Dror y sus colaboradores analizaron la totalidad de los certificados de defunción emitidos durante un período de diez años por patólogos forenses encargados de determinar si la muerte de niños pequeños había sido accidental u homicida, y hallaron evidencia de que factores ajenos a la evidencia médica, como la raza de la víctima o su

relación con el cuidador a cargo, influían de manera sistemática en la clasificación final (Dror et al., 2021). Investigaciones posteriores extendieron estos hallazgos al análisis de manchas de sangre, al reconocimiento facial asistido por tecnología, a la genética forense y a la informática forense, entre otras disciplinas (Kukucka y Dror, 2023).

Un dato revelador de la magnitud del problema surge de una encuesta global realizada entre examinadores forenses, que encontró que muchos profesionales tienen una comprensión limitada del sesgo cognitivo y tienden a considerarse a sí mismos inmunes a él, al punto de que una proporción relevante de encuestados declaró una tasa de acierto propia del cien por ciento (Sciencedirect, 2017, citado en la investigación sobre sesgo y ceguera cognitiva en peritos forenses). Esta percepción de infalibilidad es, precisamente, uno de los principales obstáculos para la adopción de medidas correctivas, ya que dificulta el reconocimiento institucional de que el sesgo no es un problema de mala fe individual, sino una característica estructural del procesamiento cognitivo humano frente a información ambigua (Dror, 2023).

Trabajos recientes han comenzado a proponer estrategias concretas de mitigación, entre ellas la exposición sistemática del examinador a hipótesis alternativas antes de emitir una conclusión definitiva, un procedimiento que ha demostrado modificar tanto el contenido de la opinión pericial como el nivel de confianza declarado por el experto (Lidén et al., 2023). Otras medidas incluyen el control estricto de la información de caso a la que accede el examinador, la verificación ciega de los resultados por parte de un segundo perito y la documentación explícita del razonamiento que sustenta cada conclusión (Kukucka y Dror, 2023). Como resume el jurista Brandon Garrett, profesor de la Facultad de Derecho de la Universidad de Duke, en poco más de una década el sesgo cognitivo pasó de ser un concepto prácticamente desconocido en el ámbito forense a convertirse en conocimiento común dentro de los laboratorios, un cambio cultural en el que la obra de Dror tuvo un papel decisivo (Garrett, 2021).

Dror (2023) propone además un modelo escalonado para clasificar el riesgo de sesgo según el tipo de tarea forense. En un extremo se ubican las tareas puramente objetivas, como el pesaje de una sustancia o la lectura de un instrumento calibrado, donde el margen para la influencia contextual es mínimo. En el otro extremo se sitúan las tareas que exigen un juicio comparativo complejo, como determinar si dos huellas dactilares latentes de baja calidad provienen de la misma persona, donde el examinador debe integrar múltiples fuentes de información ambigua y donde, en consecuencia, el riesgo de sesgo es más alto. Esta clasificación resulta especialmente relevante para el resto de este libro, porque permite anticipar en qué tipos de tareas forenses la incorporación de sistemas de inteligencia artificial podría reducir el riesgo de sesgo humano y en cuáles, por el contrario, podría simplemente desplazar ese riesgo hacia el diseño del algoritmo, hacia los datos de entrenamiento o hacia la persona encargada de interpretar su resultado (véase la Figura 1.2).

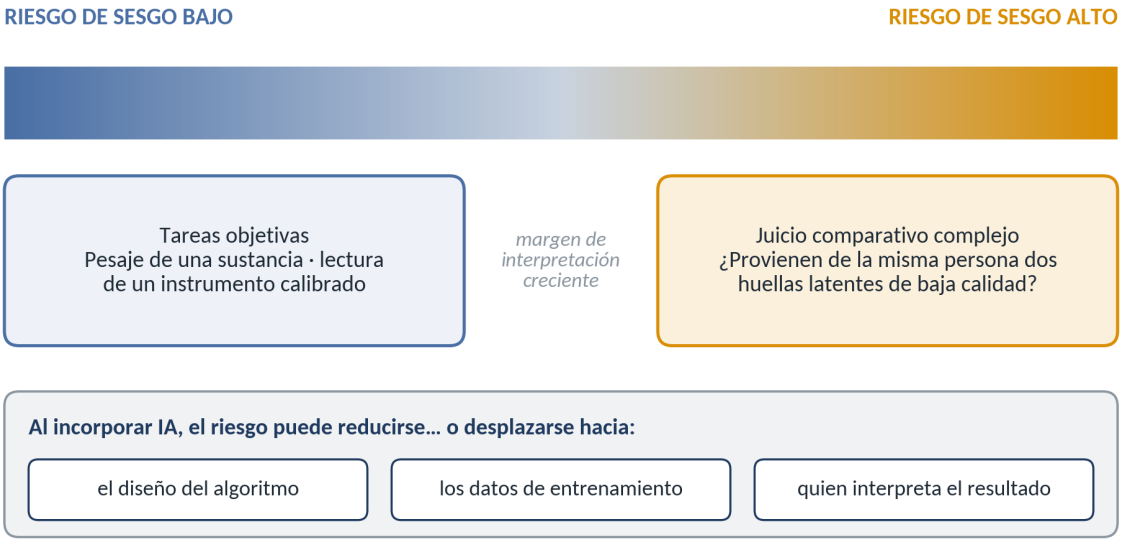


Figura 1.2. Modelo escalonado del riesgo de sesgo cognitivo según el tipo de tarea forense. El riesgo crece con el margen de interpretación; la inteligencia artificial puede reducirlo o solo desplazarlo. Fuente: elaboración propia a partir de Dror (2023).

Conviene subrayar que el reconocimiento del sesgo cognitivo no debe interpretarse como un cuestionamiento a la integridad profesional de los peritos, sino como una descripción precisa de cómo funciona la cognición humana frente a la incertidumbre. Ningún grado de formación técnica, por riguroso que sea, elimina por completo la tendencia natural del cerebro humano a buscar coherencia entre la nueva información y las expectativas previas. Precisamente por eso, las soluciones más efectivas no apelan a la fuerza de voluntad individual del examinador, sino al rediseño de los procedimientos institucionales que determinan qué información recibe, en qué momento la recibe y quién verifica de manera independiente sus conclusiones.

1.7. Respuestas institucionales: hacia una criminalística basada en evidencia

La publicación del informe de 2009 desencadenó, particularmente en Estados Unidos, un conjunto de iniciativas orientadas a fortalecer las bases científicas de la disciplina. El Departamento de Justicia y el National Institute of Standards and Technology conformaron la Organización de Comités de Áreas Científicas, conocida por su sigla en inglés OSAC, encargada de desarrollar y revisar estándares técnicos específicos para cada disciplina forense, y se creó también, con carácter temporal, la Comisión Nacional de Ciencia Forense, un organismo consultivo integrado por representantes del ámbito judicial, académico y pericial (Ballantyne y Butler, 2015).

Estas iniciativas se enmarcan en un movimiento más amplio de estandarización internacional, del cual forman parte normas como la ISO/IEC 17025, referida a la competencia técnica de laboratorios de ensayo y calibración, y la más reciente ISO 21043, específicamente diseñada para procesos de ciencia forense. La adopción de estos marcos normativos no resuelve por sí sola el problema de fondo señalado por los informes NAS y PCAST, pero establece un piso mínimo de trazabilidad, control de calidad y verificación externa que antes de 2009 era, en muchos países, prácticamente inexistente. Este capítulo volverá sobre estos estándares con mayor detalle en el apartado dedicado específicamente a la validación de herramientas forenses, incluidas aquellas basadas en inteligencia artificial.

La lógica que subyace a estas reformas institucionales puede resumirse en un principio relativamente simple, aunque de aplicación compleja: ningún método debería ingresar a la práctica

forense rutinaria, ni ser presentado ante un tribunal como prueba científica, sin haber atravesado antes un proceso de validación externo, documentado y sujeto a revisión periódica. Ese principio, evidente cuando se formula en abstracto, tropieza en la práctica con incentivos institucionales que no siempre favorecen su cumplimiento. Los laboratorios forenses suelen operar bajo presión de reducir el tiempo de respuesta ante la acumulación de casos pendientes, los presupuestos destinados a investigación metodológica son habitualmente menores que los destinados al procesamiento operativo de evidencia, y la incorporación de un nuevo método suele producirse por decisión de la propia institución que lo utilizará, sin una instancia de certificación verdaderamente independiente. Comprender estas tensiones estructurales resulta indispensable para no repetir, en la introducción de la inteligencia artificial, los mismos errores de gobernanza que explican buena parte de la crisis descrita en los apartados anteriores.

1.8. La criminalística en Iberoamérica: entre la reforma procesal y el rezago tecnológico

La transformación de la criminalística en América Latina ha estado íntimamente ligada al proceso de reforma de los sistemas de justicia penal, que a partir de la década de 1990 impulsó el reemplazo de modelos inquisitivos por sistemas de corte acusatorio en la mayoría de los países de la región (Taboga, 2022). Este cambio de paradigma procesal trasladó buena parte del peso probatorio a la etapa de juicio oral, donde la calidad, la trazabilidad y la solidez científica de la prueba pericial adquieren una relevancia central para el resultado del proceso, a diferencia de los sistemas anteriores donde la prueba se construía predominantemente durante la instrucción escrita.

En este contexto, instituciones como la Academia Mexicana de Criminalística, fundada en 1975, han desempeñado un papel relevante como referente histórico y científico de los servicios periciales en la región, al promover la formación de comunidades científicas capaces de discutir y validar el conocimiento forense emergente (Lázaro Ruiz, 2024). No obstante, persisten desafíos estructurales significativos, entre ellos la heterogeneidad en la acreditación de laboratorios, la escasez de programas de formación de posgrado especializados en criminalística y la falta de protocolos unificados entre jurisdicciones, todo lo cual dificulta la comparabilidad de resultados periciales entre países e incluso entre instituciones de un mismo país.

La distinción entre prueba pericial y prueba científica que plantea Verdú (2023) resulta especialmente pertinente para el contexto iberoamericano, donde muchos sistemas procesales todavía carecen de mecanismos equivalentes al estándar Daubert estadounidense para evaluar la admisibilidad de la evidencia científica novedosa, lo que deja en manos del juez, sin criterios técnicos explícitos, la valoración de metodologías periciales cuya solidez empírica puede variar enormemente. Este vacío normativo cobra una relevancia adicional en el momento en que la región comienza a incorporar herramientas de inteligencia artificial al trabajo forense, a menudo sin un marco de validación previo equivalente al que, con todas sus limitaciones, existe hoy en Estados Unidos o en la Unión Europea.

A esta situación se suma una dimensión de desigualdad interna dentro de la propia región. Los institutos de medicina legal y ciencias forenses de las principales capitales latinoamericanas suelen contar con equipamiento relativamente moderno y personal capacitado en técnicas actuales, mientras que las delegaciones regionales y municipales, responsables de procesar la enorme mayoría de los casos que ingresan al sistema penal, operan frecuentemente con recursos limitados, alta rotación de personal y escasa actualización tecnológica. Esta brecha interna, documentada en distintos diagnósticos sobre laboratorios forenses en la región, plantea un desafío adicional para cualquier estrategia de incorporación de inteligencia artificial: una herramienta que funcione

adecuadamente en un laboratorio central bien equipado puede resultar inaccesible, o simplemente inaplicable, en el contexto donde efectivamente se procesa la mayor parte de la evidencia del sistema.

Por otra parte, la reforma procesal penal acusatoria trajo consigo un cambio cultural que todavía está en curso en buena parte de la región: el paso de una prueba pericial escrita, incorporada al expediente sin mayor contradicción, a una prueba pericial oral, sometida al examen directo y al contrainterrogatorio de las partes en juicio. Este cambio exige que el perito no solo domine la técnica que aplica, sino que sea capaz de explicar sus fundamentos, sus limitaciones y su margen de error ante un tribunal y, eventualmente, ante un jurado. La irrupción de sistemas de inteligencia artificial añade una capa adicional de complejidad a este desafío comunicacional, porque exige que el perito pueda explicar no solo su propio razonamiento, sino también, en la medida de lo posible, el razonamiento del sistema que utilizó como apoyo, un problema conocido en la literatura especializada como el desafío de la explicabilidad algorítmica y que se abordará en detalle en el capítulo dedicado a la prueba algorítmica ante los tribunales.

1.9. El umbral de la inteligencia artificial: por qué esta transformación es distinta

Es en este escenario, marcado por una crisis de validez todavía no resuelta y por una toma de conciencia creciente sobre el papel del sesgo cognitivo, donde irrumpe la inteligencia artificial como herramienta forense. Una revisión sistemática reciente identificó un crecimiento sostenido de publicaciones sobre aplicaciones de inteligencia artificial en ciencias forenses entre 2014 y 2024, con especial concentración en genética forense, reconocimiento de imágenes y análisis de patrones (Ketsekioulafis et al., 2024). Otra revisión centrada específicamente en medicina forense documentó aplicaciones que van desde la estimación del intervalo post mortem hasta el diagnóstico asistido por imágenes, y concluyó que la inteligencia artificial funciona mejor como herramienta de apoyo que como sustituto del criterio del experto humano (Volonnino et al., 2024).

La pregunta que atraviesa el resto de este libro no es, por tanto, si la inteligencia artificial mejorará la exactitud de determinadas tareas forenses, algo que la evidencia empírica disponible sugiere de manera creciente, sino en qué condiciones puede hacerlo sin reproducir, a una escala potencialmente mucho mayor, los mismos problemas de validación insuficiente y de sesgo sistemático que los informes NAS y PCAST identificaron en los métodos forenses tradicionales. Un algoritmo entrenado con datos sesgados, evaluado sin estudios independientes de caja negra y presentado ante un tribunal como una caja negra en el sentido computacional del término, corre el riesgo de convertirse en una versión automatizada, y no necesariamente más objetiva, de los mismos defectos que esta disciplina lleva más de una década intentando corregir. Ese es precisamente el problema que exploran los capítulos siguientes.

Existe, sin embargo, una diferencia cualitativa entre el desafío de validación que enfrentaron los métodos forenses tradicionales y el que enfrentan hoy los sistemas de inteligencia artificial. Los primeros se desarrollaron de manera gradual, a lo largo de décadas, dentro de comunidades profesionales relativamente pequeñas y con una difusión geográfica lenta. Los segundos pueden desarrollarse, entrenarse, distribuirse comercialmente e implementarse en decenas de laboratorios de distintos países en el lapso de pocos años, impulsados por un mercado tecnológico global que no siempre coincide con los tiempos, ni con las prioridades, de la validación científica rigurosa. Esta aceleración obliga a repensar los mecanismos tradicionales de acreditación y control de calidad, diseñados originalmente para un ritmo de innovación mucho más lento que el que caracteriza al desarrollo actual de la inteligencia artificial. Los capítulos ocho y nueve de este libro retomarán esta tensión entre velocidad de innovación y rigor de validación, que constituye, en definitiva, el

hilo conductor de toda la obra.

A modo de cierre de este primer capítulo, conviene formular con claridad la tesis que se desarrollará a lo largo del libro. La inteligencia artificial no representa, para la criminalística, una ruptura absoluta respecto de su historia, sino la continuación, con nuevas herramientas y a una escala inédita, de un mismo proyecto que Locard inauguró hace más de un siglo: extraer del mundo material la mayor cantidad posible de información confiable sobre un hecho pasado, y hacerlo con un rigor que resista el escrutinio científico y judicial. La diferencia decisiva no está en la naturaleza de ese proyecto, que permanece intacta, sino en la responsabilidad que recae sobre quienes lo llevan adelante en esta nueva etapa, obligados a aprender de la crisis de validez descrita en este capítulo para no reproducirla, esta vez, bajo la forma de un algoritmo.

1.10. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos más relevantes que se desprenden del recorrido histórico presentado en este capítulo se encuentran la persistencia de métodos forenses no validados empíricamente, la subestimación institucional del sesgo cognitivo como fuente sistemática de error, la ausencia de estándares de admisibilidad probatoria suficientemente exigentes en buena parte de los sistemas procesales latinoamericanos, y el riesgo de trasladar estos mismos déficits de validación a las nuevas herramientas basadas en inteligencia artificial sin someterlas a un escrutinio equivalente al exigido para los métodos tradicionales.

Buenas prácticas recomendadas

La experiencia acumulada desde 2009 sugiere un conjunto de buenas prácticas que resultan aplicables tanto a los métodos forenses clásicos como a las herramientas emergentes de inteligencia artificial. Entre ellas destacan la exigencia de estudios independientes de caja negra antes de incorporar cualquier método a la práctica pericial rutinaria, el control estricto de la información contextual irrelevante a la que accede el examinador, la implementación de mecanismos de verificación ciega, la documentación explícita y trazable del razonamiento pericial, y la adopción de estándares internacionales de calidad como referencia mínima de acreditación institucional.

A estas prácticas conviene añadir una recomendación de carácter más general, dirigida especialmente a los responsables institucionales de laboratorios y servicios periciales en la región iberoamericana: ninguna decisión de incorporar una nueva tecnología forense, sea esta un software de comparación biométrica, un sistema de reconstrucción tridimensional de escenas o un algoritmo de genotipado probabilístico, debería adoptarse exclusivamente sobre la base de la promesa comercial del proveedor. La evidencia recogida en este capítulo demuestra que la brecha entre lo que un método promete y lo que efectivamente puede demostrar mediante estudios independientes ha sido, históricamente, la principal fuente de errores judiciales atribuibles a la ciencia forense. Exigir esa evidencia antes de adoptar una herramienta, y no después de que haya comenzado a producir resultados presentados ante un tribunal, es quizás la lección más importante que este primer capítulo deja para el resto del libro.



CAPÍTULO 2

Inteligencia Artificial para el forense: fundamentos

2.1. Caso de apertura: cuando el algoritmo también se equivoca

En diciembre de 2019, el National Institute of Standards and Technology de Estados Unidos publicó el tercer informe de su programa de evaluación de proveedores de reconocimiento facial, un estudio que analizó cerca de doscientos algoritmos desarrollados por casi cien empresas distintas, aplicados sobre más de dieciocho millones de imágenes correspondientes a más de ocho millones de personas (Grother et al., 2019). El hallazgo central del informe resultó inquietante para cualquier institución que estuviera considerando incorporar esta tecnología a su trabajo forense: la inmensa mayoría de los algoritmos evaluados mostró diferencias sistemáticas de exactitud según el sexo, la edad y el grupo étnico de las personas fotografiadas, con tasas de falso positivo que en algunos casos variaban entre diez y más de cien veces según el grupo demográfico analizado (Grother et al., 2019).

Este hallazgo no proviene de un crítico externo de la tecnología, sino de la propia agencia científica del gobierno estadounidense encargada de fijar estándares de medición, y constituye probablemente el estudio más riguroso y de mayor escala jamás realizado sobre el desempeño diferencial de los sistemas de reconocimiento facial. Su relevancia para este capítulo es directa: ilustra, con una precisión que ningún argumento abstracto podría igualar, que un sistema de inteligencia artificial no es una herramienta neutral que simplemente reemplaza el juicio humano por un cálculo objetivo. Es, más bien, un artefacto construido a partir de decisiones humanas, entrenado con datos que reflejan el mundo tal como fue registrado, y cuyo comportamiento puede heredar, amplificar o incluso introducir nuevas formas de error que un examinador sin formación técnica adecuada difícilmente podrá anticipar.

Comprender por qué ocurre esto, y no simplemente que ocurre, es el propósito de este segundo capítulo. Antes de analizar, en los capítulos siguientes, las aplicaciones concretas de la inteligencia artificial en cada disciplina forense, resulta indispensable construir un vocabulario conceptual común que permita a peritos, fiscales, jueces y defensores formular las preguntas correctas frente a cualquier herramienta algorítmica que se les presente como prueba científica.

Vale la pena subrayar, además, que el caso del reconocimiento facial no es una excepción aislada dentro del panorama de la inteligencia artificial aplicada a tareas de alto riesgo. Estudios independientes realizados por equipos académicos, y no solo por la propia agencia gubernamental responsable de los estándares, replicaron hallazgos similares con algoritmos comerciales y de código abierto, y documentaron que el desempeño diferencial no se limita al reconocimiento facial sino que aparece también, con distinta magnitud, en otras tareas de visión computacional aplicadas a rostros y cuerpos humanos. Este patrón de replicación independiente es, en sí mismo, una buena noticia desde el punto de vista científico, porque confirma que el problema es real y medible, y no un artefacto de un único estudio o de un único conjunto de datos. Es también, precisamente por eso, una advertencia que ninguna institución forense debería subestimar antes de incorporar este tipo de tecnología a sus procedimientos de identificación.

2.2. Qué entendemos por inteligencia artificial en el contexto forense

El término inteligencia artificial se utiliza con frecuencia como si designara una tecnología única y homogénea, cuando en realidad describe un campo amplio de la informática dedicado a construir sistemas capaces de realizar tareas que, tradicionalmente, requerían inteligencia humana, como reconocer un rostro, interpretar un texto o predecir un resultado a partir de datos incompletos (Russell y Norvig, 2021). Dentro de ese campo amplio conviene distinguir tres niveles conceptuales que se emplearán a lo largo de todo este libro y que rara vez son sinónimos, aunque el lenguaje periodístico y, en ocasiones, el propio lenguaje judicial tienden a confundirlos.

El primer nivel es la inteligencia artificial propiamente dicha, entendida como el objetivo general de dotar a una máquina de capacidades cognitivas. El segundo nivel es el aprendizaje automático, un enfoque particular para alcanzar ese objetivo que consiste en que el sistema aprenda patrones directamente a partir de datos de ejemplo, en lugar de que un programador le indique de manera explícita, regla por regla, cómo debe comportarse (Colliot, 2023). El tercer nivel es el aprendizaje profundo, una familia específica de técnicas de aprendizaje automático basada en redes neuronales artificiales de múltiples capas, particularmente eficaz para procesar datos complejos como imágenes, audio o texto (Janiesch et al., 2021). Estos tres niveles no son sinónimos, sino conjuntos anidados: todo aprendizaje profundo es aprendizaje automático, y todo aprendizaje automático es una forma de inteligencia artificial, pero la relación inversa no se cumple (véase la Figura 2.1).

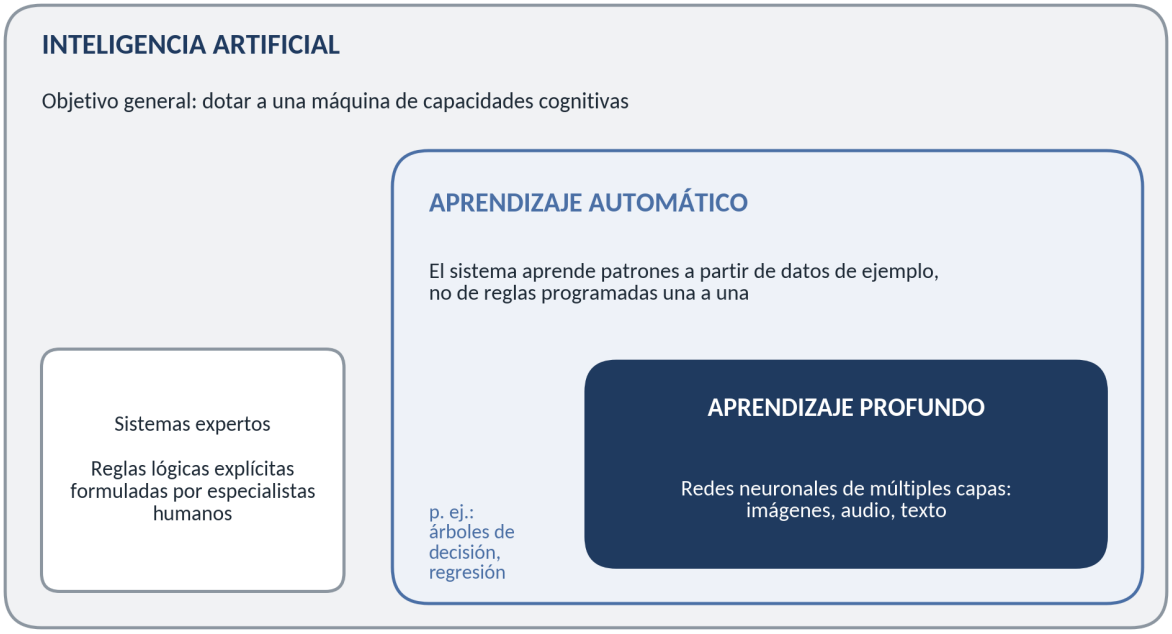


Figura 2.1. Relación anidada entre inteligencia artificial, aprendizaje automático y aprendizaje profundo. Todo aprendizaje profundo es aprendizaje automático, pero no a la inversa; los sistemas expertos son inteligencia artificial sin aprendizaje a partir de datos. Fuente: elaboración propia a partir de Russell y Norvig (2021), Colliot (2023) y Janiesch et al. (2021).

Conviene añadir, para completar este mapa conceptual, que buena parte de la historia de la inteligencia artificial anterior al auge actual del aprendizaje automático estuvo dominada por los llamados sistemas expertos, programas basados en reglas lógicas explícitas formuladas directamente por especialistas humanos del dominio correspondiente. Estos sistemas tuvieron aplicaciones tempranas en medicina y en algunas tareas forenses puntuales, y presentaban la ventaja de ser plenamente transparentes en su razonamiento, ya que cada conclusión podía rastrearse hasta la regla específica que la había generado. Su desventaja, que explica en gran medida por qué el campo se desplazó hacia el aprendizaje automático a partir de la década de 2010, era su fragilidad frente a situaciones no contempladas explícitamente por sus diseñadores y el enorme esfuerzo manual que exigía codificar el conocimiento experto en forma de reglas. Esta tensión histórica entre transparencia y flexibilidad reaparece, casi sin variaciones, en el debate contemporáneo sobre explicabilidad que se desarrolla más adelante en este mismo capítulo.

Esta distinción, aparentemente técnica, tiene consecuencias prácticas directas para el trabajo pericial. Cuando un proveedor comercial anuncia que su producto utiliza inteligencia artificial, esa afirmación no dice nada, por sí sola, sobre el tipo específico de algoritmo empleado, sobre

la cantidad y calidad de los datos con los que fue entrenado, ni sobre su grado de transparencia interna. Exigir precisión terminológica a los proveedores de tecnología forense no es un capricho académico, sino un primer paso indispensable para poder formular preguntas de validación adecuadas, en línea con las exigencias de foundational validity que plantearon los informes NAS y PCAST analizados en el capítulo anterior.

2.3. Aprendizaje automático: conceptos fundamentales

El aprendizaje automático se organiza habitualmente en tres grandes paradigmas. El aprendizaje supervisado entrena un modelo a partir de ejemplos ya etiquetados por un humano, de modo que el sistema aprende a predecir una categoría o un valor numérico a partir de nuevos datos de entrada; este es, con diferencia, el paradigma más utilizado en aplicaciones forenses actuales, como la clasificación de una imagen entre alterada o auténtica, o la estimación de la edad de un individuo a partir de una radiografía (Colliot, 2023). El aprendizaje no supervisado, en cambio, busca estructuras o agrupaciones ocultas en los datos sin contar con etiquetas previas, una técnica útil para detectar anomalías o para agrupar indicios similares dentro de una gran base de datos de casos. El aprendizaje por refuerzo, finalmente, entrena a un sistema mediante un proceso de prueba y error guiado por recompensas, y aunque su aplicación directa a la criminalística todavía es limitada, comienza a explorarse en tareas de planificación de rutas de investigación y de optimización de recursos periciales.

Todo sistema de aprendizaje supervisado se construye a partir de tres elementos que cualquier perito debería poder identificar antes de confiar en sus resultados. El primero son los datos de entrenamiento, es decir, el conjunto de ejemplos etiquetados a partir de los cuales el sistema aprende sus patrones; la calidad, el tamaño y la representatividad de ese conjunto determinan, en gran medida, el desempeño posterior del modelo. El segundo son las características o variables que el sistema utiliza para describir cada ejemplo, que pueden ser definidas manualmente por un experto, un proceso conocido como ingeniería de características, o extraídas automáticamente por el propio modelo a partir de los datos crudos, como ocurre en la mayoría de los sistemas de aprendizaje profundo (Colliot, 2023). El tercero es la función de pérdida, una medida matemática del error que el algoritmo intenta minimizar progresivamente durante el proceso de entrenamiento.

Un concepto central para evaluar cualquier sistema forense basado en aprendizaje automático es la distinción entre el conjunto de entrenamiento y el conjunto de prueba. Un modelo debe evaluarse siempre con datos que no haya visto durante su entrenamiento, porque de lo contrario se corre el riesgo de sobreajuste, un fenómeno en el cual el sistema memoriza las particularidades del conjunto de entrenamiento en lugar de aprender patrones generalizables, y por lo tanto rinde de manera excelente en las pruebas internas del desarrollador, pero de manera mucho más pobre cuando se enfrenta a casos reales, distintos de aquellos con los que fue construido. Este fenómeno explica, en parte, por qué un algoritmo puede alcanzar cifras de exactitud extraordinarias en la publicación técnica de su fabricante y, sin embargo, comportarse de manera mucho menos confiable en el laboratorio que lo adopta, un patrón que recuerda directamente a la advertencia central del informe PCAST sobre la necesidad de estudios independientes de caja negra, realizados por evaluadores ajenos al desarrollador del método.

Otro concepto que el perito forense debe manejar con soltura es el de métricas de desempeño, porque un mismo modelo puede describirse de maneras muy distintas según qué métrica se elija para presentarlo. La exactitud general, que mide simplemente la proporción de predicciones correctas, resulta engañosa cuando las categorías que el sistema intenta distinguir están muy desbalanceadas, como ocurre, por ejemplo, cuando se busca detectar contenido manipulado entre un volumen

enorme de contenido auténtico. En esos casos resulta más informativo examinar por separado la sensibilidad del sistema, es decir, su capacidad de detectar correctamente los casos positivos, y su especificidad, su capacidad de no generar falsas alarmas sobre casos negativos, junto con la matriz de confusión completa que desglosa los cuatro tipos posibles de resultado: verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Como se señaló en el capítulo anterior a propósito del sesgo hacia los falsos positivos en los propios informes NAS y PCAST, cualquier evaluación de un sistema forense, humano o algorítmico, que solo reporte una de estas dos tasas de error ofrece una imagen incompleta y potencialmente engañosa de su desempeño real.

2.4. Aprendizaje profundo y redes neuronales

El aprendizaje profundo se apoya en redes neuronales artificiales, estructuras matemáticas inspiradas de manera muy general en el funcionamiento del cerebro humano, organizadas en capas de unidades interconectadas que transforman progresivamente la información de entrada hasta producir una salida (Janiesch et al., 2021). La característica que distingue al aprendizaje profundo del aprendizaje automático tradicional es la profundidad de estas redes, es decir, el número de capas intermedias que el dato atraviesa antes de llegar a la salida final, lo que permite al sistema construir representaciones progresivamente más abstractas de la información, desde bordes y texturas simples en las primeras capas de una imagen hasta conceptos complejos como rostros u objetos completos en las capas finales.

Dentro del aprendizaje profundo existen arquitecturas especializadas para distintos tipos de datos. Las redes neuronales convolucionales se diseñaron específicamente para procesar imágenes, al explotar la estructura espacial de los píxeles mediante filtros que detectan patrones locales, y constituyen hoy la base de la mayoría de las herramientas forenses de análisis de imagen (Kim y Lee, 2023). Las redes recurrentes y, más recientemente, la arquitectura conocida como transformador, se especializan en el procesamiento de datos secuenciales como el texto o el habla, y son el fundamento técnico tanto de los sistemas de procesamiento del lenguaje natural utilizados en lingüística forense computacional como de los grandes modelos de lenguaje que sustentan la actual generación de inteligencia artificial generativa (véase la Figura 2.2).

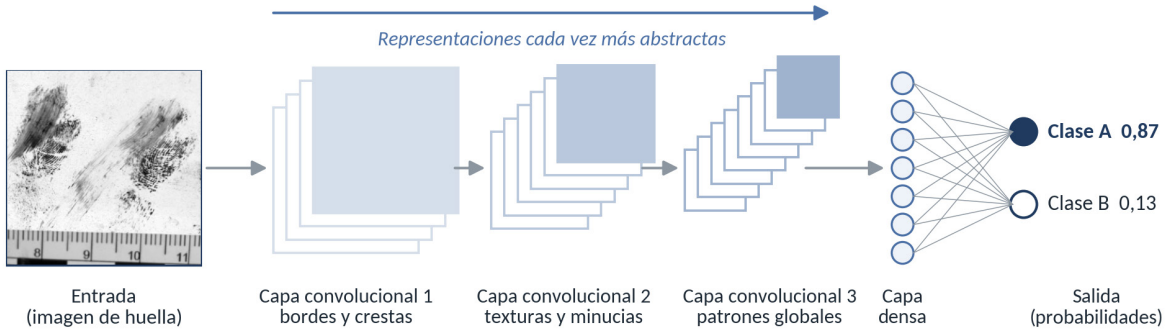


Figura 2.2. Esquema simplificado de una red neuronal convolucional aplicada a una imagen de huella latente. Cada capa extrae rasgos más abstractos que la anterior hasta producir probabilidades de clase; las cifras son ilustrativas. Fuente: elaboración propia; imagen de entrada del archivo del autor, perito forense Freddy Rodolfo Sánchez Polit.

Una limitación estructural del aprendizaje profundo, con consecuencias directas para su uso forense, es que su capacidad predictiva depende de manera crítica del volumen y la diversidad de los datos de entrenamiento disponibles. A diferencia de otras áreas donde existen enormes bases de datos públicas disponibles para entrenar modelos, buena parte de la evidencia forense es escasa, sensible desde el punto de vista de la privacidad, y en muchos casos legalmente restringida en

su acceso, lo que limita el tamaño de los conjuntos de datos disponibles para entrenar y validar modelos específicos de cada disciplina forense. Esta escasez de datos ayuda a explicar por qué, como se documentó en el capítulo anterior, buena parte de la literatura científica sobre inteligencia artificial forense proviene de un número reducido de países y poblaciones, principalmente de Asia oriental, lo que introduce interrogantes legítimos sobre la generalización de esos hallazgos a poblaciones distintas de las estudiadas (Ketsekioulafis et al., 2024).

Esta observación tiene una implicación directa y todavía poco discutida para la región iberoamericana. Un modelo de aprendizaje profundo entrenado predominantemente con imágenes dentales, radiografías óseas o fotografías faciales de poblaciones asiáticas o europeas no puede asumirse válido, sin más, para su aplicación sobre poblaciones latinoamericanas, caracterizadas por un grado de mestizaje genético y una diversidad fenotípica considerablemente distintos de los conjuntos de datos con los que fue entrenado originalmente. La exigencia de estudios de validación local, adaptados a las poblaciones efectivamente atendidas por cada sistema de justicia, no es entonces un requisito burocrático adicional, sino una condición científica mínima para que cualquier herramienta de este tipo pueda considerarse forense y no meramente experimental cuando se aplica fuera del contexto donde fue desarrollada.

2.5. Visión computacional aplicada a la evidencia forense

La visión computacional es la rama de la inteligencia artificial dedicada a que las máquinas extraigan información significativa a partir de imágenes o videos, y constituye, en la práctica, el área de aplicación más desarrollada dentro de la ciencia forense actual. Sus tareas típicas incluyen la clasificación de una imagen completa en una categoría determinada, la detección de objetos específicos dentro de una imagen, la segmentación precisa de regiones de interés y la comparación biométrica entre dos o más imágenes de un mismo sujeto potencial (Volonnino et al., 2024).

En el ámbito de la identificación humana, la visión computacional se aplica hoy a tareas tan diversas como la estimación de la edad a partir de radiografías dentales o de la muñeca, la estimación del sexo biológico a partir de medidas craneométricas, la reconstrucción facial aproximada a partir de restos óseos y, de manera particularmente extendida, la comparación de rostros entre una imagen cuestionada y una base de datos de referencia (Ketsekioulafis et al., 2024). Este último uso, el reconocimiento facial, es también el que concentra la mayor controversia, precisamente por los hallazgos de desempeño diferencial documentados en el caso de apertura de este capítulo.

Resulta importante que el perito forense distinga con precisión entre dos tareas que el lenguaje cotidiano suele confundir bajo la etiqueta genérica de reconocimiento facial. La verificación uno a uno consiste en confirmar si dos imágenes específicas corresponden a la misma persona, una tarea relativamente más sencilla y con tasas de error generalmente más bajas. La identificación uno a muchos, en cambio, consiste en buscar coincidencias de una imagen cuestionada dentro de una base de datos potencialmente enorme, una tarea considerablemente más compleja, con un margen de error mayor y con consecuencias mucho más serias en caso de un falso positivo, ya que puede señalar como sospechosa a una persona completamente ajena al hecho investigado (Grother et al., 2019). Los capítulos siguientes de este libro, en particular el dedicado a la identificación humana aumentada, retomarán esta distinción con mayor detalle técnico.

2.6. Procesamiento del lenguaje natural y lingüística forense computacional

El procesamiento del lenguaje natural agrupa las técnicas de inteligencia artificial dedicadas a que una máquina interprete, genere o analice texto y habla humana. Su aplicación forense más

consolidada es la lingüística forense computacional, un campo que estudia rasgos de estilo, vocabulario y estructura sintáctica con el fin de estimar la probabilidad de que un texto cuestionado haya sido producido por un autor determinado, o de perfilar características del autor de un texto anónimo, como su procedencia geográfica o su lengua materna (Alduais et al., 2023).

El fundamento teórico de esta disciplina es la noción de idiolecto, la idea de que cada persona posee un patrón relativamente estable y distintivo de uso del lenguaje que se manifiesta de manera consistente a través de sus producciones orales y escritas, de forma análoga, aunque nunca idéntica en su grado de certeza científica, a la huella dactilar o al perfil genético (Nini, 2023). Durante las últimas dos décadas, la disciplina desarrolló métodos computacionales cada vez más sofisticados para medir ese idiolecto de manera cuantitativa, mediante técnicas de estilometría que analizan la frecuencia de palabras funcionales, la longitud media de las oraciones o patrones sintácticos característicos de cada autor.

La irrupción de los grandes modelos de lenguaje ha introducido una tensión completamente nueva en este campo. Por un lado, estos modelos ofrecen herramientas de análisis de corpus mucho más potentes y escalables que las disponibles anteriormente, capaces de procesar volúmenes de texto que ningún examinador humano podría abordar manualmente. Por otro lado, esos mismos modelos permiten que cualquier persona module, imite o enmascare su propio estilo de escritura con una facilidad sin precedentes, lo que erosiona precisamente el supuesto de estabilidad del idiolecto sobre el cual se ha construido buena parte de la disciplina durante décadas (Mikros, 2025). Investigaciones recientes muestran, además, que los sistemas actuales de detección de texto generado por inteligencia artificial presentan tasas elevadas de falsos positivos, particularmente cuando se aplican a textos escritos por personas cuya lengua materna no es el inglés, y resultan vulnerables a estrategias adversariales relativamente sencillas de evasión (Mikros, 2025). Esta doble condición, herramienta poderosa y amenaza simultánea a los fundamentos metodológicos de la propia disciplina, ilustra con particular claridad la tensión que atraviesa a toda la criminalística contemporánea frente a la inteligencia artificial.

Trabajos experimentales recientes exploran, además, el uso de grandes modelos de lenguaje para tareas de perfilado de autor en textos digitales, con resultados que muestran un potencial considerable para automatizar tareas de clasificación de estilo, pero que también advierten sobre las limitaciones persistentes de estos modelos frente a estilos de escritura deliberadamente disfrazados u obfuscados, un escenario común en comunicaciones amenazantes o en textos vinculados a actividad delictiva (Cho et al., 2024). Esta limitación refuerza un principio que atraviesa todo este capítulo: la inteligencia artificial amplía de manera notable la capacidad analítica del perito, pero no elimina la necesidad de un juicio experto capaz de contextualizar e interpretar sus resultados.

Un ejemplo concreto de aplicación forense del procesamiento del lenguaje natural, más allá de la atribución de autoría propiamente dicha, es el análisis automatizado de amenazas y comunicaciones extorsivas, donde los sistemas actuales pueden clasificar de manera preliminar un texto según su nivel de riesgo, identificar patrones lingüísticos asociados a determinados perfiles de emisor, o detectar indicios de coacción en declaraciones escritas. Estas aplicaciones resultan particularmente valiosas en contextos de investigación con grandes volúmenes de comunicaciones digitales, como los que se generan en causas de extorsión, trata de personas o delincuencia organizada, donde el volumen de mensajes a analizar excede ampliamente la capacidad de revisión manual de cualquier equipo de investigadores. Sin embargo, la literatura especializada coincide en que estas clasificaciones automatizadas deben entenderse siempre como una herramienta de priorización y apoyo a la investigación, y no como una conclusión pericial definitiva sobre la autoría o la intención de un texto determinado, distinción que resulta crucial para evitar que un indicio orientador se

transforme, sin la validación adecuada, en prueba de cargo.

2.7. Inteligencia artificial generativa: creación y detección de contenido sintético

La inteligencia artificial generativa agrupa a los sistemas capaces de crear contenido nuevo, texto, imágenes, audio o video, a partir de patrones aprendidos de grandes volúmenes de datos, en lugar de limitarse a clasificar o analizar contenido ya existente. Sus arquitecturas más influyentes incluyen las redes generativas antagónicas, en las cuales dos redes neuronales compiten entre sí, una generando contenido sintético y otra intentando distinguirlo del contenido real, y los grandes modelos de lenguaje basados en la arquitectura de transformador, capaces de generar texto coherente a partir de instrucciones en lenguaje natural (Pei et al., 2024).

Esta misma tecnología que permite crear contenido sintético de manera cada vez más realista es la que sustenta el fenómeno de los deepfakes, imágenes, audios o videos manipulados o generados artificialmente que representan de manera falsa a una persona diciendo o haciendo algo que nunca ocurrió. La proliferación de este tipo de contenido ha crecido de manera muy acelerada en los últimos años, con estimaciones que documentan un incremento del orden de cuatro veces en la cantidad de deepfakes detectados en esquemas de fraude entre 2023 y 2024 (Deepfake-Eval-2024, citado en el conjunto de datos correspondiente). Investigaciones recientes muestran, además, que la capacidad de las personas comunes para distinguir contenido auténtico de contenido sintético se ha reducido de manera significativa conforme mejora la calidad de los generadores (Deepfake-Eval-2024).

Frente a este escenario, la propia comunidad de visión computacional ha desarrollado técnicas forenses de detección de contenido sintético, que analizan artefactos estadísticos sutiles dejados por el proceso de generación, como inconsistencias en el dominio de frecuencia, irregularidades en el parpadeo ocular o incoherencias fisiológicas imperceptibles para el ojo humano (Pei et al., 2024). Sin embargo, esta relación entre generación y detección tiene una característica estructural que la distingue de cualquier otra disciplina forense analizada en este libro: se trata de una carrera armamentista continua, en la cual cada avance en las técnicas de detección incentiva el desarrollo de nuevas técnicas de generación diseñadas específicamente para evadir esos detectores, lo que hace que ningún sistema de detección de contenido sintético pueda considerarse válido de manera permanente, a diferencia de un método de comparación de huellas dactilares, cuyos fundamentos físicos no cambian con el tiempo (véase la Figura 2.3).

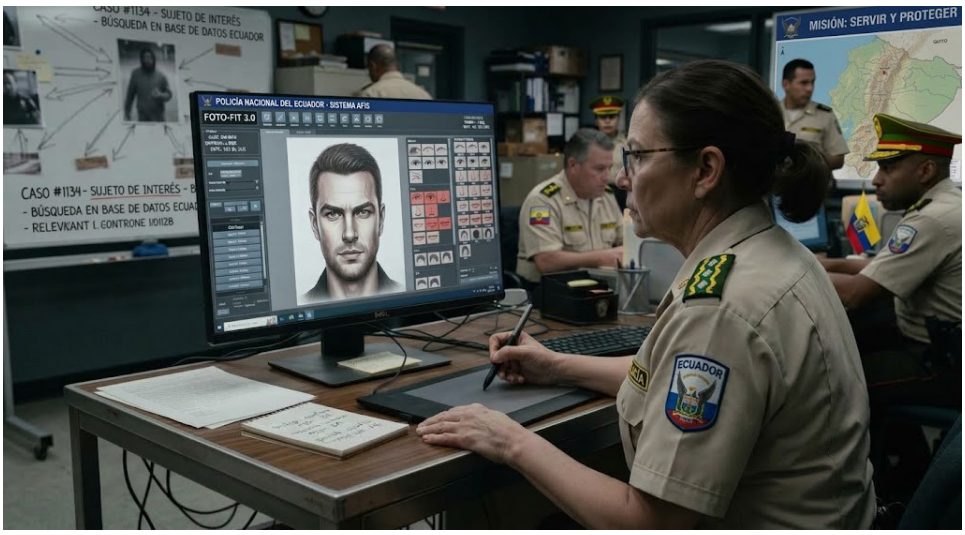


Figura 2.3. Imagen sintética de una oficina de identificación con software de retrato compuesto. Obsérvense los indicios de generación artificial: textos sin sentido en la pizarra («RELEVKANT I.

CONTRONE»), una interfaz rotulada como «Sistema AFIS» que en realidad muestra un programa de retrato hablado, e insignias y rótulos que no corresponden a ningún modelo real. Ilustración generada con inteligencia artificial con fines didácticos; no corresponde a evidencia real. No representa a ninguna institución ni a personas reales.

Un desarrollo particularmente relevante para el trabajo forense es la exploración de los propios grandes modelos de lenguaje multimodales como herramientas de detección de contenido sintético, al solicitarles que evalúen directamente si una imagen determinada fue generada artificialmente. Los resultados preliminares de estas investigaciones son mixtos: estos modelos muestran cierta capacidad para identificar indicios de manipulación, pero también exhiben limitaciones significativas de consistencia y son susceptibles a errores que resultan difíciles de anticipar sin una evaluación empírica exhaustiva (Jia et al., 2024, citado en Pei et al., 2024). Esta incertidumbre técnica reafirma la necesidad, planteada ya en el capítulo anterior, de que ninguna herramienta de este tipo se incorpore a la práctica pericial sin haber sido sometida antes a estudios de validación independientes.

Conviene además distinguir con claridad entre dos escenarios de uso que la práctica forense suele confundir bajo la etiqueta genérica de detección de deepfakes. El primer escenario es la autenticación de un archivo específico que llega a manos del investigador, donde la pregunta es si ese archivo concreto fue alterado o generado artificialmente, una tarea en la que las técnicas forenses tradicionales de análisis de metadatos y de artefactos de compresión siguen siendo, en muchos casos, más robustas y mejor validadas que los detectores basados exclusivamente en aprendizaje profundo. El segundo escenario es la vigilancia continua de grandes volúmenes de contenido circulante en redes sociales o plataformas digitales, donde la escala del problema exige necesariamente herramientas automatizadas, pero donde la tasa de error tolerable debe calibrarse de manera muy distinta a la de un dictamen pericial individual, porque el costo de un error aislado en un sistema de monitoreo masivo no es equivalente al costo de un error en la prueba que sustenta una acusación penal concreta.

2.8. El problema de la caja negra: explicabilidad e interpretabilidad

Uno de los desafíos más consecuentes que plantea la inteligencia artificial para el trabajo forense es lo que la literatura especializada denomina el problema de la caja negra: muchos de los modelos más precisos, en particular las redes neuronales profundas, generan sus predicciones mediante cálculos matemáticos tan complejos que resulta extremadamente difícil, incluso para sus propios desarrolladores, explicar en términos comprensibles por qué el sistema llegó a una conclusión determinada (Brożek et al., 2023). Esta opacidad entra en tensión directa con una exigencia central del proceso penal, según la cual toda prueba admitida en juicio debe poder ser explicada, cuestionada y sometida a contradicción por las partes.

Frente a este desafío ha surgido el campo de la inteligencia artificial explicable, dedicado a desarrollar técnicas que permitan comprender, al menos de manera aproximada, el razonamiento interno de un modelo complejo. Estas técnicas se dividen habitualmente en dos grandes familias. Los métodos intrínsecamente interpretables utilizan modelos cuya propia estructura es transparente, como los árboles de decisión, que sacrifican en ocasiones algo de exactitud predictiva a cambio de una interpretabilidad directa. Los métodos de explicación posterior, en cambio, intentan aproximar el comportamiento de un modelo opaco ya entrenado mediante técnicas como LIME o SHAP, que estiman la contribución de cada variable de entrada a una predicción específica, sin modificar el modelo original (Explainable AI and Law, 2023).

Conviene señalar que existe un debate técnico activo, y todavía no resuelto, sobre si la disyuntiva entre exactitud predictiva e interpretabilidad es realmente inevitable. Un sector influyente de la investigación en aprendizaje automático sostiene que, para la mayoría de los problemas de alto riesgo, incluidos los del sistema de justicia penal, existen modelos intrínsecamente interpretables capaces de alcanzar una exactitud comparable a la de los modelos de caja negra, lo que convertiría la elección de un modelo opaco en una decisión de conveniencia técnica más que en una necesidad científica insalvable. Desde esta perspectiva, la explicabilidad no debería entenderse como una limitación que hay que tolerar a cambio de mayor precisión, sino como un requisito de diseño que puede y debe exigirse desde el inicio del desarrollo de cualquier herramienta destinada a producir prueba forense.

En el ámbito específico de la informática forense digital, distintos estudios recientes advierten que las herramientas de inteligencia artificial actualmente disponibles todavía no satisfacen plenamente los estándares de trazabilidad y defensibilidad legal que exige un proceso penal, y proponen marcos de trabajo que integran explicabilidad desde el diseño inicial de la herramienta, en lugar de añadirla como una capa posterior (Billah, 2025). Esta misma idea, la explicabilidad como requisito de diseño y no como parche posterior, reaparecerá de manera recurrente en el capítulo de este libro dedicado específicamente a la admisibilidad de la prueba algorítmica ante los tribunales.

Resulta útil, para el lector no especializado en informática, distinguir entre dos preguntas que suelen confundirse al hablar de explicabilidad. La primera pregunta es técnica: qué variables de entrada influyeron más en una predicción concreta, algo que herramientas como SHAP pueden estimar con razonable precisión matemática. La segunda pregunta es epistémica, y resulta mucho más exigente: por qué esas variables constituyen, en el caso concreto, una razón válida y suficiente para sostener una conclusión pericial ante un tribunal. Una técnica de explicabilidad puede responder satisfactoriamente a la primera pregunta sin resolver en absoluto la segunda, porque identificar qué píxeles de una imagen influyeron en la clasificación de un algoritmo no equivale, todavía, a poder justificar esa clasificación con el mismo estándar argumentativo que exige el derecho probatorio. Esta brecha entre explicación técnica y justificación jurídica constituye, en la actualidad, uno de los terrenos de investigación más activos en la intersección entre inteligencia artificial y ciencias forenses (véase la Figura 2.4).

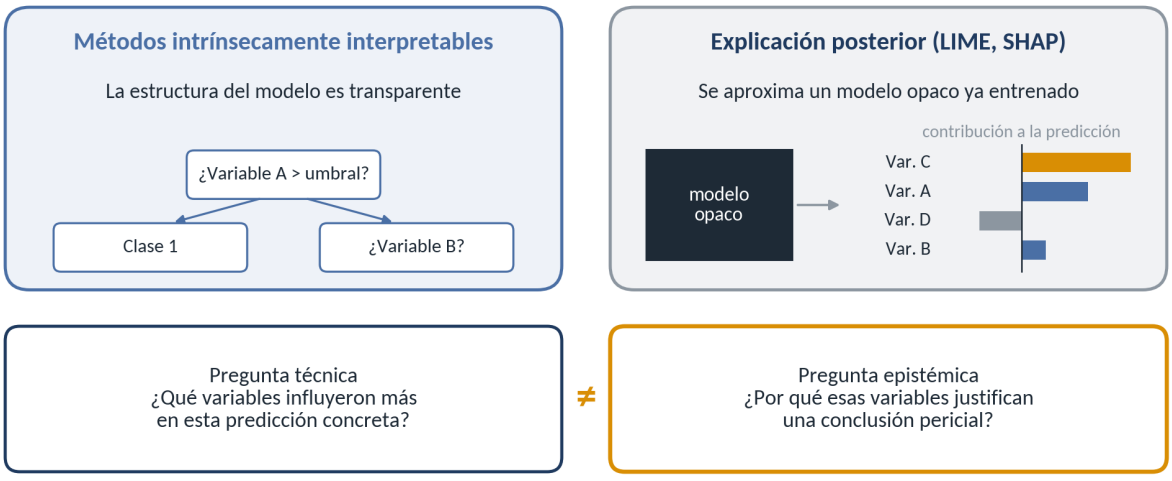


Figura 2.4. Las dos familias de técnicas de inteligencia artificial explicable y las dos preguntas que suelen confundirse. Identificar qué variables influyeron en una predicción no equivale a justificar por qué sostienen una conclusión pericial. Fuente: elaboración propia a partir de Explainable AI and Law (2023).

2.9. Qué puede y qué no puede hacer hoy un algoritmo con evidencia forense

A modo de síntesis de este capítulo, conviene precisar con claridad los límites actuales de la inteligencia artificial aplicada al trabajo forense, para evitar tanto el entusiasmo acrítico como el rechazo igualmente acrítico frente a esta tecnología. Un sistema de inteligencia artificial puede hoy procesar volúmenes de datos muy superiores a los que cualquier examinador humano podría abordar, detectar patrones sutiles que escapan a la percepción humana no asistida, aplicar el mismo criterio de manera perfectamente consistente a miles de comparaciones sucesivas sin fatiga ni variabilidad, y ofrecer estimaciones cuantitativas de incertidumbre allí donde el juicio humano tradicionalmente solo ofrecía una opinión categórica.

Un sistema de inteligencia artificial no puede, en cambio, garantizar por sí mismo que sus resultados sean válidos fuera del conjunto de datos con el que fue entrenado, explicar de manera completa y siempre satisfactoria el razonamiento que condujo a una conclusión específica, corregir automáticamente los sesgos presentes en sus datos de entrenamiento sin una intervención humana deliberada, ni sustituir la responsabilidad legal y ética que recae, en última instancia, sobre el perito humano que firma un dictamen. Los grandes modelos de lenguaje, en particular, presentan además una limitación adicional bien documentada en la literatura técnica reciente: la tendencia a generar información que suena plausible y está formulada con fluidez, pero que resulta objetivamente incorrecta o directamente inventada, un fenómeno conocido como alucinación y que constituye, hoy por hoy, una limitación estructural de estos sistemas y no simplemente un defecto que se corregirá con futuras versiones (Ji et al., 2023; Xu et al., 2024).

Esta síntesis conceptual proporciona el vocabulario y el criterio necesarios para analizar, en los capítulos siguientes, cómo se aplican concretamente estas tecnologías en cada disciplina forense específica, desde la identificación humana hasta la genética, pasando por el procesamiento de la escena del crimen y el análisis de rastros materiales, sin perder de vista, en ningún momento, que la pregunta relevante nunca es únicamente qué puede hacer un algoritmo, sino qué puede demostrar de manera científicamente válida y qué puede explicar de manera legalmente satisfactoria.

Cabe cerrar este capítulo con una observación que conecta directamente con la tesis general de este libro, ya anticipada al final del capítulo anterior. La inteligencia artificial no introduce, en rigor, categorías de problemas completamente nuevas para la criminalística. Introduce, en cambio, una versión acelerada y a mayor escala de problemas que la disciplina ya conocía: la necesidad de validar empíricamente cualquier método antes de utilizarlo como prueba, la necesidad de conocer y comunicar su tasa de error, y la necesidad de que un ser humano capacitado pueda explicar, ante un tribunal, por qué una conclusión determinada merece ser creída. Lo que cambia con la inteligencia artificial no es la naturaleza de estas exigencias, sino la velocidad con la que las nuevas herramientas se desarrollan y se ofrecen comercialmente, y la sofisticación matemática que ahora se requiere para satisfacer esas mismas exigencias de siempre. Los capítulos que siguen, dedicados a las aplicaciones específicas de esta tecnología en cada disciplina forense, deben leerse siempre con este marco conceptual como telón de fondo.

2.10. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la confusión terminológica que dificulta exigir estándares de validación adecuados a los proveedores de tecnología, el desempeño diferencial de numerosos sistemas de visión computacional según variables demográficas, la erosión de supuestos metodológicos tradicionales de disciplinas como

la lingüística forense frente a la inteligencia artificial generativa, la opacidad de los modelos de mayor exactitud predictiva y su difícil compatibilidad con las exigencias de explicación propias del proceso penal, y la tendencia de los grandes modelos de lenguaje a producir información incorrecta con apariencia de plausibilidad.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir a cualquier proveedor de tecnología forense una descripción precisa del tipo de algoritmo empleado y de las características de sus datos de entrenamiento, solicitar siempre estudios de desempeño diferencial por subgrupo demográfico antes de adoptar un sistema biométrico, priorizar, cuando la exactitud sea comparable, modelos intrínsecamente interpretables por sobre modelos de caja negra, verificar de manera independiente cualquier resultado producido por inteligencia artificial generativa antes de incorporarlo a un dictamen pericial, y mantener siempre un criterio experto humano como responsable final de toda conclusión forense, sin importar cuán sofisticada sea la herramienta algorítmica utilizada como apoyo.

A estas prácticas conviene sumar una recomendación de carácter formativo, dirigida a las instituciones responsables de la enseñanza de la criminalística en la región. La formación de nuevos peritos no puede seguir asumiendo que el dominio de una disciplina forense tradicional es suficiente para evaluar críticamente una herramienta de inteligencia artificial aplicada a esa misma disciplina. Comprender la diferencia entre un conjunto de entrenamiento y un conjunto de prueba, entre sensibilidad y especificidad, o entre una explicación técnica y una justificación jurídica, ya no es un conocimiento optativo reservado a especialistas en informática, sino un componente cada vez más central de la competencia profesional exigible a cualquier perito que trabaje, directa o indirectamente, con evidencia procesada mediante estas tecnologías. El capítulo dedicado al perito del futuro, hacia el final de este libro, retomará esta cuestión formativa con mayor detalle.

CAPÍTULO 3

Identificación humana aumentada: biometría, reconocimiento facial, AFIS con IA, antropología y odontología forense

3.1. Caso de apertura: el caso Mayfield y los límites de la certeza dactiloscópica

El 11 de marzo de 2004, una serie de atentados con bombas en trenes de cercanías de Madrid causó casi doscientas muertes y más de mil cuatrocientos heridos. Entre la evidencia recuperada por la policía española se encontró una bolsa de plástico con detonadores, sobre la cual se detectó una huella dactilar latente. Esa huella fue remitida al FBI, cuyo laboratorio, mediante su sistema automatizado de identificación dactilar, generó una lista de veinte candidatos con características compatibles. Tres examinadores del FBI, y posteriormente un perito independiente designado por la corte, concluyeron con lo que describieron como absoluta certeza que la huella correspondía a Brandon Mayfield, un abogado de Portland, Oregón, sin antecedentes penales, converso al islam y que nunca había estado en España (Departamento de Justicia de Estados Unidos, Oficina del Inspector General, 2006).

Mayfield fue arrestado como testigo material y permaneció detenido durante dos semanas, mientras el FBI autorizaba vigilancia encubierta sobre su domicilio y su despacho. La policía española, sin embargo, insistió repetidamente ante el FBI en que la huella no correspondía a Mayfield, hasta que finalmente identificó a un ciudadano argelino como su verdadero origen. El FBI retiró su identificación, se disculpó formalmente y liberó a Mayfield. La investigación posterior del inspector general del Departamento de Justicia concluyó que el error no se debió a una falla del sistema automatizado, que había cumplido correctamente su función de generar una lista de candidatos plausibles, sino a un exceso de confianza de los examinadores humanos, quienes permitieron que la coincidencia de aproximadamente diez puntos característicos, una similitud estadísticamente extraordinaria pero no concluyente, se transformara en una identificación categórica, y que además ignoraron de manera reiterada las advertencias de un organismo par igualmente calificado (Departamento de Justicia de Estados Unidos, Oficina del Inspector General, 2006).

El caso Mayfield resulta especialmente valioso para abrir este capítulo porque desmiente, con un ejemplo documentado por una investigación oficial exhaustiva, dos mitos que todavía circulan sobre la identificación biométrica. El primero es que el sistema automatizado, por sí solo, fue el origen del error, cuando en realidad el algoritmo cumplió su función limitada de proponer candidatos y fue el juicio humano posterior el que convirtió una similitud notable en una certeza infundada. El segundo es que la dactiloscopia constituye una ciencia de individualización absoluta, sin margen de error, un supuesto que el propio caso contribuyó a desmontar y que autores como Cole (2009) han analizado en profundidad desde la epistemología de la identificación forense. Ambas lecciones resultan decisivas para comprender, en las páginas siguientes, qué cambia y qué no cambia cuando la inteligencia artificial se incorpora a las distintas disciplinas dedicadas a la identificación humana (véase la Figura 3.1).



Figura 3.1. Huellas latentes parciales y superpuestas, reveladas con polvo sobre una superficie oscura.

Obsérvense las crestas interrumpidas, los arrastres y el ruido de fondo que reducen la información útil: el tipo de evidencia que, como en el caso Mayfield, exige extrema cautela antes de afirmar una identificación. Fuente: Archivo del autor, perito forense Freddy Rodolfo Sánchez Polit.

Conviene añadir un tercer elemento del caso, menos citado que los dos anteriores pero igualmente relevante para este libro: la investigación del inspector general documentó que los examinadores del FBI, una vez formada su convicción inicial, procesaron toda la información posterior de manera que confirmara esa convicción, incluidas las reiteradas objeciones de la policía española, en lugar de tratarlas como una señal de alarma que ameritaba una revisión completa del análisis. Este patrón coincide de manera casi exacta con el mecanismo de sesgo de confirmación descrito por Dror y sus colaboradores y analizado en el primer capítulo de este libro, lo que confirma que el caso Mayfield no fue únicamente un fallo técnico aislado, sino una manifestación temprana y de altísimo perfil público de exactamente el mismo problema estructural que, años más tarde, los informes NAS y PCAST documentarían de manera sistemática para el conjunto de la ciencia forense estadounidense.

3.2. Qué entendemos por identificación humana aumentada

La identificación humana forense reúne el conjunto de disciplinas dedicadas a determinar quién es una persona, viva o fallecida, a partir de evidencia física recuperada en una escena, en un cuerpo o en restos óseos. Este capítulo aborda cuatro de esas disciplinas, la biometría dactilar, el reconocimiento facial, la antropología forense y la odontología forense, unidas por un mismo patrón de transformación: la incorporación de sistemas de aprendizaje profundo que automatizan tareas que, hasta hace pocos años, dependían por completo del juicio visual y la experiencia acumulada de un examinador humano (véase la Figura 3.2).



Figura 3.2. Cuerpo en examen externo, con indicadores de evidencia. En personas fallecidas, la identificación humana comienza con la documentación sistemática del cuerpo, antes de cualquier análisis asistido por algoritmos. Imagen editada: el rostro y las zonas íntimas se pixelaron para proteger la dignidad y la identidad de la persona fallecida. Fuente: Archivo del autor, perito forense Freddy Rodolfo Sánchez Polit.

El adjetivo aumentada, en el título de este capítulo, se elige de manera deliberada en lugar de automatizada o reemplazada. Como se anticipó en el capítulo anterior, la evidencia disponible sugiere de manera consistente que los sistemas de inteligencia artificial rinden mejor cuando

funcionan como una herramienta que extiende la capacidad analítica del perito, y no cuando se conciben como un sustituto autónomo de su criterio. Este capítulo documenta, disciplina por disciplina, en qué medida esa promesa de aumento se cumple en la práctica actual, y en qué medida persisten limitaciones que exigen mantener al examinador humano en el centro del proceso de decisión final.

Las cuatro disciplinas analizadas en este capítulo comparten, además, un desafío común que conviene enunciar desde el inicio: todas ellas dependen, en mayor o menor medida, de comparar un rasgo biológico individual contra un patrón de referencia poblacional, ya sea una base de datos de huellas, un banco de imágenes faciales, una colección osteológica de referencia o un conjunto de radiografías dentales de edad conocida. La calidad de esa comparación depende tanto de la calidad de la muestra cuestionada, frecuentemente parcial, deteriorada o incompleta en el contexto forense real, como de la representatividad del patrón de referencia contra el cual se compara. Mantener presente esta doble dependencia, de la muestra y del patrón, ayudará a interpretar con criterio crítico las cifras de exactitud que se presentan a lo largo de este capítulo, casi siempre obtenidas en condiciones de laboratorio más favorables que las que enfrenta un perito en un caso real.

Conviene, finalmente, anticipar la lógica expositiva que sigue este capítulo. Cada uno de los cuatro apartados centrales presenta primero el estado del arte técnico de la disciplina correspondiente, con las cifras de desempeño que reporta la literatura más reciente, y examina después, de manera igualmente detallada, los límites conocidos de esa misma tecnología. Esta estructura deliberadamente simétrica responde a una convicción que atraviesa todo este libro: presentar únicamente las cifras de exactitud favorables de una tecnología, sin sus correspondientes limitaciones documentadas, constituye una forma de desinformación tan perjudicial para la práctica forense responsable como el rechazo acrítico de cualquier innovación tecnológica.

3.3. Biometría dactilar: de AFIS al aprendizaje profundo

Los sistemas automatizados de identificación dactilar, conocidos por su sigla en inglés AFIS, existen desde la década de 1960 y constituyen la aplicación biométrica más antigua y consolidada de la informática a la investigación criminal. Su funcionamiento tradicional se basa en la extracción de minucias, puntos característicos de la cresta papilar como bifurcaciones y terminaciones, que se comparan de manera algorítmica contra una base de datos de referencia para generar una lista corta de candidatos, cuya decisión final de identificación queda siempre en manos de un examinador humano (Herke, 2026) (véase la Figura 3.3).



Figura 3.3. Crestas papilares de la yema de un dedo. Las bifurcaciones y terminaciones de estas crestas son las minucias que los sistemas AFIS extraen y comparan. Fuente: Fredy Rodolfo Sánchez Polit.

La distinción más importante para comprender el estado actual de esta disciplina es la que separa las huellas impresas, tomadas de manera controlada mediante tinta o escáner directo sobre los dedos de una persona, de las huellas latentes, recuperadas de manera incidental de una superficie en la escena del crimen. Para las primeras, los sistemas AFIS actuales alcanzan niveles de automatización y exactitud muy altos, con un desempeño que se acerca a la operación completamente autónoma. Para las segundas, la automatización completa sigue siendo un desafío no resuelto, debido a la calidad frecuentemente deficiente de la huella latente, la superposición con ruido de fondo y el área parcial habitualmente disponible para el análisis (Herke, 2026) (véase la Figura 3.4).

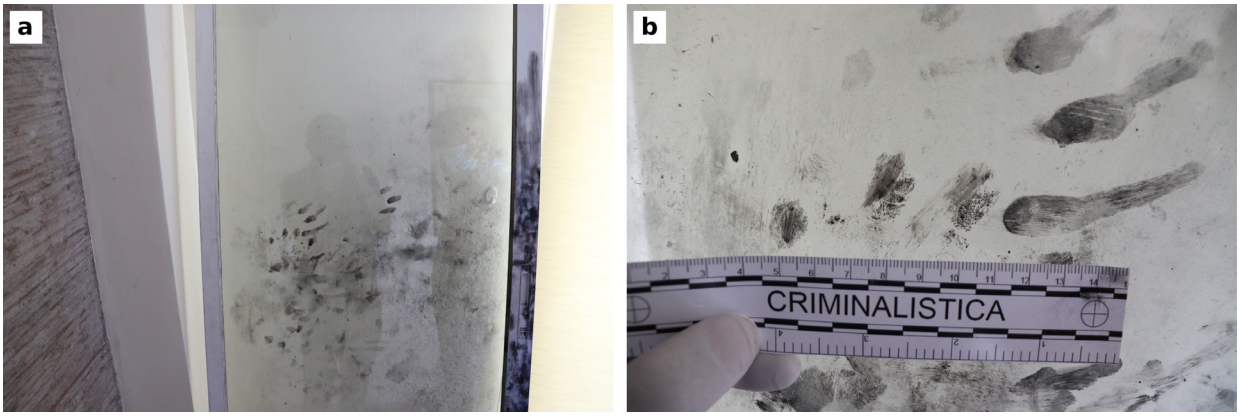


Figura 3.4. Revelado de huellas latentes con polvo sobre una puerta de vidrio: a) vista general del soporte; b) acercamiento con testigo métrico. Obsérvese la calidad variable de las latentes, con arrastres y superposiciones, frente a la de una huella impresa tomada en condiciones controladas. Fuente: Fredy Rodolfo Sánchez Polit.

Esta brecha de desempeño entre huellas impresas y huellas latentes explica por qué la mayor parte del flujo de trabajo pericial en dactiloscopia sigue organizándose, incluso en los laboratorios mejor equipados, en torno al llamado modelo de niveles crecientes de automatización. En el nivel más básico, el sistema únicamente asiste en la digitalización y el almacenamiento de las imágenes. En un nivel intermedio, el sistema propone candidatos ordenados por similitud, pero requiere que un examinador marque manualmente las minucias antes de ejecutar la búsqueda. En el nivel más avanzado, todavía poco extendido en la práctica policial cotidiana de la región, el sistema extrae las minucias de manera completamente automática y presenta directamente una lista de candidatos ordenada, reservando la intervención humana para la etapa final de verificación. Conocer en qué nivel de este continuo opera efectivamente el sistema AFIS de una institución determinada es, para cualquier perito, un requisito mínimo antes de poder explicar con precisión, ante un tribunal, qué papel jugó exactamente el algoritmo y qué papel jugó el juicio humano en una identificación concreta.

La incorporación de redes neuronales convolucionales ha producido avances medibles en este terreno históricamente difícil. Sistemas recientes basados en arquitecturas de atención profunda logran fusionar información local, correspondiente a las minucias tradicionales, con representaciones globales de la textura de la cresta papilar, lo que mejora de manera sustancial la exactitud de coincidencia sobre huellas latentes de baja calidad respecto de los métodos anteriores basados exclusivamente en minucias (Grosz y Jain, 2023). Estos avances resultan significativos porque, como documentó la evaluación independiente del NIST sobre reconocimiento latente, incluso los mejores sistemas automatizados disponibles hace una década apenas alcanzaban una tasa de identificación correcta en el primer lugar de la lista de candidatos de alrededor del setenta por ciento, una cifra que subraya por qué la revisión humana sigue siendo indispensable en esta tarea.

Conviene subrayar, sin embargo, que ningún avance técnico en la etapa de comparación algorítmica resuelve por sí solo la pregunta epistemológica de fondo que el caso Mayfield puso sobre la mesa: cuántos puntos de coincidencia son suficientes para sostener una identificación individual, y con qué grado de certeza puede formularse esa conclusión. Cole (2009) argumenta que la dactiloscopia, pese a su prestigio histórico, nunca contó con una base estadística equivalente a la de la genética forense para cuantificar la probabilidad de una coincidencia casual, un vacío que la incorporación de inteligencia artificial no elimina automáticamente, salvo que los propios sistemas se diseñen desde el inicio para expresar sus resultados en términos probabilísticos explícitos, con un margen de error conocido, en lugar de reproducir el lenguaje categórico de certeza absoluta que caracterizó tradicionalmente a esta disciplina.

Un desarrollo institucional relevante para la región iberoamericana es la creciente interoperabilidad entre sistemas AFIS de distintos países y organismos, que permite compartir consultas biométricas a través de fronteras nacionales de manera cada vez más ágil. Esta interoperabilidad, impulsada en buena medida por acuerdos regionales de cooperación policial y migratoria, ofrece ventajas evidentes para la investigación de delitos transnacionales, pero también multiplica las consecuencias potenciales de cualquier error de identificación no corregido, ya que una conclusión errónea generada en un país puede propagarse automáticamente a las bases de datos y a los procesos de investigación de otro, sin que necesariamente medie una nueva revisión humana independiente en cada jurisdicción receptora. Este riesgo de propagación transfronteriza del error refuerza la necesidad, ya planteada en el capítulo anterior, de que cada sistema AFIS documente con claridad su tasa de error y las condiciones bajo las cuales fue validado, información que debería viajar junto con cada resultado compartido entre agencias.

Finalmente, conviene mencionar que la propia comunidad técnica dedicada al desarrollo de sistemas AFIS reconoce, de manera explícita, los límites actuales de la automatización plena. Los manuales de referencia más autorizados de la disciplina insisten en que, incluso en los sistemas más avanzados basados en aprendizaje profundo, la etapa final de verificación de una coincidencia propuesta por el algoritmo debe quedar siempre a cargo de un examinador certificado, siguiendo protocolos de revisión ciega que impidan que el propio resultado numérico del sistema condicione, por un efecto de anclaje, el juicio independiente del perito humano (Maltoni et al., 2022). Este principio de verificación ciega, que retomará con mayor detalle el capítulo de este libro dedicado a la validación de herramientas forenses, constituye hoy el consenso profesional dominante en biometría dactilar, incluso entre los desarrolladores de la tecnología más avanzada (véase la Figura 3.5).

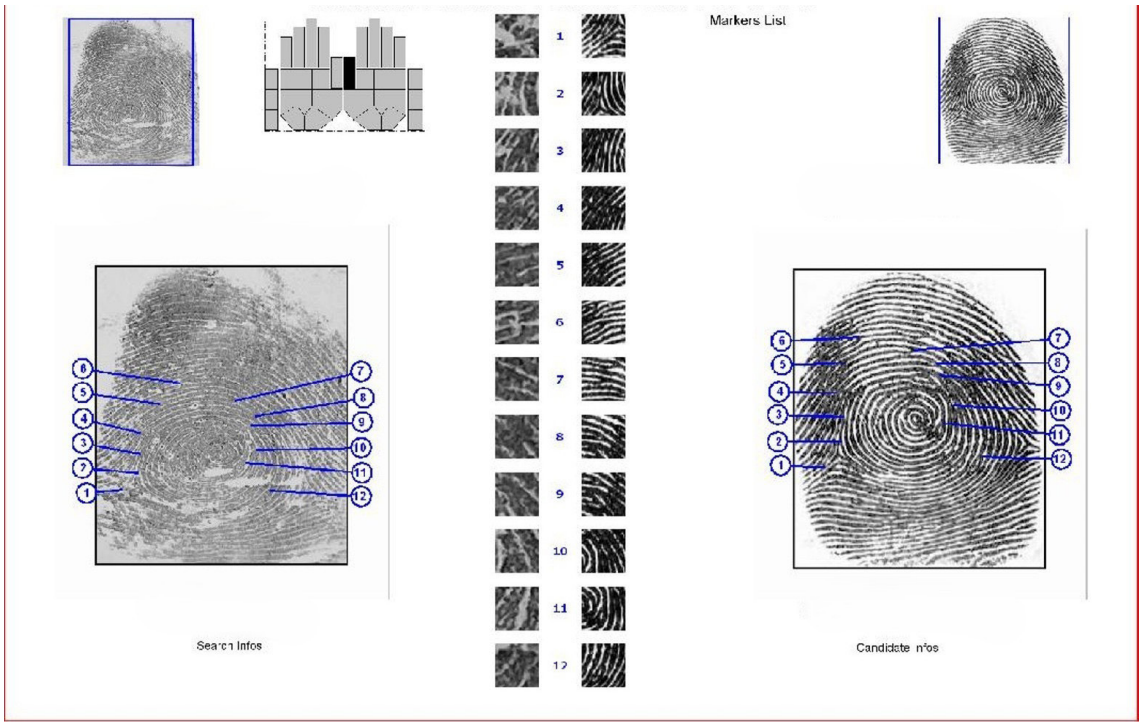


Figura 3.5. Pantalla de comparación de un sistema AFIS con doce puntos característicos correlacionados. A la izquierda, la huella dubitada; a la derecha, la del candidato propuesto. El sistema propone; la verificación final corresponde al examinador. Fuente: Fredy Rodolfo Sánchez Polit.

3.4. Reconocimiento facial forense: verificación, identificación y sus límites

El reconocimiento facial mediante aprendizaje profundo se ha convertido, en poco más de una década, en una de las aplicaciones de mayor difusión de la inteligencia artificial forense, impulsada por la disponibilidad masiva de cámaras de videovigilancia y de bases de datos fotográficas gubernamentales. Como se explicó en el capítulo anterior a propósito del informe del NIST sobre efectos demográficos, resulta indispensable distinguir entre la verificación uno a uno, que confirma si dos imágenes corresponden a la misma persona, y la identificación uno a muchos, que busca coincidencias de una imagen cuestionada dentro de una base de datos completa (Grother et al., 2019). Esta segunda tarea, la más habitual en investigaciones criminales que buscan identificar a una persona desconocida a partir de una grabación de video, es también la que concentra el mayor riesgo de error y la mayor controversia social.

El hallazgo central del informe del NIST, ya expuesto como caso de apertura del capítulo anterior, merece retomarse aquí en su aplicación específica al contexto de la identificación criminal: las diferencias de exactitud según sexo, edad y grupo étnico no son un defecto menor y corregible con ajustes superficiales, sino una característica estructural de la mayoría de los algoritmos comerciales evaluados, con tasas de falso positivo que varían, según el grupo demográfico, por factores de diez a más de cien veces (Grother et al., 2019). Trasladado a un escenario de investigación criminal real, esto significa que un sistema de reconocimiento facial aplicado sobre una base de datos poblacional puede generar, para ciertos grupos demográficos, un número de falsas coincidencias considerablemente mayor que para otros, con el consiguiente riesgo de que personas completamente ajenas a un hecho delictivo sean señaladas como sospechosas con mayor frecuencia según su pertenencia a un grupo determinado.

Esta evidencia ha llevado a que numerosas guías profesionales y marcos regulatorios recientes

insistan en que ningún resultado de un sistema de reconocimiento facial debería considerarse, por sí solo, una identificación positiva, sino únicamente una pista investigativa que requiere corroboración independiente mediante otros métodos, incluida la revisión visual cuidadosa de un examinador capacitado y, siempre que sea posible, evidencia adicional no derivada del propio sistema algorítmico. Esta recomendación, coherente con el principio general expuesto en el capítulo dos sobre el papel del algoritmo como apoyo y no como sustituto del criterio experto, resulta particularmente urgente en el caso del reconocimiento facial dada la magnitud y la sistematicidad del desempeño diferencial documentado.

Un aspecto adicional que el perito y el operador judicial deben considerar es la enorme diferencia de condiciones entre las imágenes con las que se validan comercialmente estos sistemas y las imágenes que efectivamente se procesan en una investigación criminal real. Las cifras de exactitud más favorables que ofrecen los desarrolladores suelen obtenerse sobre fotografías frontales, bien iluminadas y de alta resolución, condiciones que rara vez coinciden con las de una grabación de videovigilancia captada a distancia, con ángulo oblicuo, baja resolución y condiciones de iluminación deficientes, el escenario más habitual en la práctica de investigación criminal. Esta brecha entre el desempeño de laboratorio y el desempeño de campo, documentada de manera reiterada en la literatura técnica sobre reconocimiento facial, constituye una razón adicional, independiente del problema del desempeño demográfico diferencial, para tratar con extrema cautela cualquier resultado de identificación facial obtenido a partir de material audiovisual de calidad subóptima.

Por último, conviene señalar que la controversia sobre el reconocimiento facial forense no se limita a su exactitud técnica, sino que se extiende a un debate más amplio sobre vigilancia masiva y derechos fundamentales, particularmente cuando estos sistemas se despliegan de manera continua sobre espacios públicos y no únicamente como herramienta puntual de investigación de un hecho ya ocurrido. Este capítulo se concentra en la dimensión técnica y probatoria del problema, pero el capítulo de este libro dedicado a sesgo algorítmico, privacidad y derechos humanos retomará esta discusión desde una perspectiva normativa más amplia, indispensable para cualquier marco regulatorio que pretenda gobernar el uso de esta tecnología de manera responsable.

3.5. Antropología forense asistida: estimación de sexo y edad mediante inteligencia artificial

La antropología forense se ocupa de la identificación de restos óseos, particularmente relevante en contextos de desastres masivos, conflictos armados, búsqueda de personas desaparecidas y restos esqueletizados o en avanzado estado de descomposición. Dos de sus tareas centrales, la estimación del sexo biológico y la estimación de la edad al momento de la muerte, han sido objeto de una intensa actividad de investigación en inteligencia artificial durante los últimos años.

En materia de estimación de sexo, los métodos tradicionales de la disciplina se basan en la evaluación morfológica visual de rasgos del cráneo y de la pelvis por parte de un antropólogo entrenado, un proceso que, pese a su validez reconocida, está sujeto a variabilidad entre observadores y entre observaciones repetidas del mismo examinador (Wang et al., 2024). Los primeros estudios que aplicaron redes neuronales artificiales a esta tarea, utilizando imágenes tridimensionales de cráneos, reportaron niveles de exactitud de alrededor del noventa y cinco por ciento, superiores a los alcanzados por antropólogos forenses experimentados sobre el mismo conjunto de casos (Bewes et al., 2019). Estudios más recientes, aplicados a otras estructuras óseas como el hueso coxal, reportan niveles de exactitud todavía más altos, cercanos al cien por ciento, cuando se emplean arquitecturas de aprendizaje profundo entrenadas sobre reconstrucciones tridimensionales

obtenidas por tomografía computarizada (Epain et al., 2024).

Una revisión sistemática reciente sobre técnicas de estimación de sexo a partir del cráneo confirma esta tendencia general, al mostrar que los enfoques de aprendizaje profundo tienden a superar tanto a los métodos estadísticos clásicos como a los métodos de aprendizaje automático más simples, especialmente cuando se dispone de datos de imagen tridimensional de buena calidad (Wang et al., 2024). La misma revisión advierte, sin embargo, que la inmensa mayoría de los estudios disponibles se basa en muestras de una única población, frecuentemente asiática o europea, lo que limita severamente la capacidad de generalizar estos hallazgos a poblaciones con una composición genética y morfológica distinta, una advertencia que conecta directamente con la discusión sobre representatividad de los datos de entrenamiento planteada en el capítulo anterior y que adquiere una relevancia particular para las poblaciones mestizas de América Latina.

Esta limitación de representatividad poblacional no es un detalle menor para la práctica forense de la región. Los rasgos morfológicos que distinguen estadísticamente el cráneo o la pelvis masculina de la femenina varían de manera documentada entre poblaciones con distinta ascendencia genética, y un modelo entrenado predominantemente sobre una población de referencia china o europea puede desplazar sistemáticamente sus predicciones cuando se aplica sobre una población con un grado de mestizaje considerable, como ocurre en la mayoría de los países latinoamericanos. La solución a este problema no pasa por descartar estas herramientas, cuyo potencial de mejora sobre los métodos morfológicos tradicionales es real y está bien documentado, sino por exigir, antes de su adopción institucional, estudios locales de recalibración que ajusten el modelo, o que al menos midan explícitamente su desempeño, sobre una muestra de referencia representativa de la población efectivamente atendida por cada servicio forense.

En materia de estimación de edad, la investigación se ha concentrado especialmente en radiografías de mano y muñeca, útiles tanto para la determinación de la mayoría de edad legal en migrantes indocumentados o en menores no acompañados como para la identificación de víctimas en contextos de desastre masivo. Los sistemas más recientes integran, en una arquitectura de dos etapas, un primer módulo de localización anatómica automática de puntos de referencia óseos y un segundo módulo de estimación conjunta de edad y sexo, diseñados específicamente para funcionar de manera independiente de la presencia de tejido blando, una condición frecuente en restos humanos en avanzado estado de descomposición o esqueletizados. Estos sistemas alcanzan niveles de precisión comparables a los de un examinador experto, con la ventaja adicional de una interpretabilidad anatómica que facilita su defensa ante un tribunal.

La determinación de la mayoría de edad legal merece una consideración aparte, porque sus consecuencias jurídicas son particularmente significativas: distingue, según el sistema procesal de que se trate, entre un régimen penal juvenil y un régimen penal de adultos, entre la aplicación o no de determinadas medidas de protección migratoria, y entre distintos niveles de responsabilidad civil y penal. Precisamente por esa razón, la literatura especializada insiste en que ningún sistema de inteligencia artificial debería utilizarse para fijar una edad puntual y exacta en estos casos, sino para estimar un rango de edad probable con su correspondiente intervalo de incertidumbre, dejando a criterio del operador jurídico, y no del algoritmo, la decisión final sobre cómo tratar los casos que se ubican en la frontera legal entre la minoría y la mayoría de edad. Esta distinción entre estimación probabilística y determinación categórica, que atraviesa todo este capítulo, adquiere aquí una dimensión de derechos humanos que ningún avance técnico debería oscurecer.

3.6. Odontología forense e inteligencia artificial: entre la estimación de edad y la cautela sobre las marcas de mordida

La odontología forense aplica el conocimiento sobre el desarrollo y el estado de la dentición humana a dos tareas principales: la estimación de la edad, particularmente valiosa en menores y adolescentes cuyo desarrollo dental sigue una secuencia relativamente predecible, y la identificación individual mediante el cotejo de registros dentales o, de manera mucho más controvertida, mediante el análisis de marcas de mordida.

La estimación dental de la edad ha experimentado una transformación considerable con la incorporación de redes neuronales convolucionales aplicadas a radiografías panorámicas. Estudios recientes que emplean arquitecturas de transferencia de aprendizaje, entrenadas inicialmente sobre grandes bases de datos de imágenes generales y luego ajustadas específicamente sobre imágenes dentales, reportan errores absolutos medios de apenas tres años sobre poblaciones que abarcan desde los ocho hasta los sesenta y ocho años de edad (Ataş et al., 2023). Otros estudios centrados específicamente en la determinación de la mayoría de edad legal, una de las aplicaciones forenses más sensibles de esta técnica por sus implicancias en procesos migratorios y penales de menores, reportan niveles de precisión superiores al ochenta por ciento utilizando el desarrollo del tercer molar como principal indicador radiográfico (Altındağ et al., 2023).

Investigaciones más recientes, orientadas específicamente a poblaciones pediátricas y adolescentes, han desarrollado modelos capaces de estimar simultáneamente la edad y el sexo biológico a partir de una única radiografía panorámica, con un error medio inferior a un año y medio para la estimación de edad y una exactitud superior al ochenta y siete por ciento para la clasificación de sexo, aunque con un desempeño que se deteriora de manera notable en los tramos etarios más cercanos a la edad adulta, precisamente el rango de mayor relevancia legal (Sributsayakarn et al., 2025). Esta variación del desempeño según el tramo de edad ilustra un patrón que se repite en toda la odontología forense computacional: la exactitud de estos sistemas no es uniforme, sino que depende de manera crítica de qué tan cerca se encuentre el caso concreto del rango etario mejor representado en los datos de entrenamiento, una consideración que cualquier perito debe verificar antes de aplicar un modelo a un caso individual.

Desde el punto de vista metodológico, resulta valioso que el perito comprenda por qué estos modelos suelen entrenarse mediante la técnica de transferencia de aprendizaje mencionada en el capítulo anterior, en lugar de construirse desde cero para cada aplicación forense específica. Dado que las bases de datos de radiografías dentales etiquetadas con edad cronológica confirmada son, comparativamente, mucho más pequeñas que las bases de datos de imágenes generales disponibles para el entrenamiento inicial de una red neuronal, los desarrolladores aprovechan el conocimiento visual general ya aprendido por el modelo en esa primera fase, y lo ajustan después con el conjunto, más reducido, de imágenes dentales específicas. Esta estrategia explica, en parte, por qué los resultados obtenidos en poblaciones bien representadas en los estudios publicados, como la turca, la coreana o la tailandesa, no garantizan automáticamente un desempeño equivalente sobre una población forense latinoamericana, cuya composición genética y cuyos patrones de desarrollo dental pueden diferir de manera no trivial de los conjuntos de datos originales.

El panorama es radicalmente distinto cuando se trata del análisis de marcas de mordida como método de identificación individual. Como se documentó en el primer capítulo de este libro, el informe PCAST de 2016 concluyó que esta técnica carecía por completo de estudios que respaldaran su validez científica, debido a la deformación impredecible que sufre la piel humana al recibir una mordida y a la ausencia de evidencia sólida sobre la unicidad de la dentición humana (President's Council of Advisors on Science and Technology, 2016). La incorporación de inteligencia artificial

a esta tarea específica no resuelve ese problema de fondo, porque ningún algoritmo, por sofisticado que sea, puede extraer de una imagen distorsionada una información que la propia fuente material nunca contuvo con la precisión suficiente. Cualquier iniciativa futura que proponga automatizar el análisis de marcas de mordida debería, por tanto, comenzar por resolver el problema de validación científica subyacente, y no simplemente añadir una capa de automatización algorítmica sobre un método cuyos fundamentos siguen sin demostrarse.

Este contraste entre la estimación de edad dental, un terreno donde la inteligencia artificial ha producido mejoras medibles y bien documentadas, y el análisis de marcas de mordida, un terreno donde ningún grado de sofisticación algorítmica puede compensar la ausencia de un fundamento científico validado, ilustra con particular claridad uno de los mensajes centrales de este libro. La pregunta relevante frente a cualquier nueva herramienta de inteligencia artificial forense no es si la tecnología es lo suficientemente avanzada, sino si la tarea subyacente que se le pide realizar descansa sobre un fundamento científico sólido. Cuando ese fundamento existe, como ocurre con la relación bien documentada entre el desarrollo dental y la edad cronológica, la inteligencia artificial puede aportar mejoras sustanciales de objetividad y exactitud. Cuando ese fundamento no existe, como ocurre con la supuesta unicidad forense de las marcas de mordida, ninguna tecnología, por avanzada que sea, puede suplir esa ausencia.

3.7. Fusión multimodal y bases de datos biométricas interoperables

Un desarrollo reciente que atraviesa transversalmente las disciplinas descritas en este capítulo es la fusión multimodal, la combinación simultánea de distintas señales biométricas, huella dactilar, rostro, iris, voz e incluso patrones de tatuajes, dentro de una misma plataforma de identificación, con el fin de compensar las debilidades de cada modalidad individual mediante la información complementaria que aportan las demás. Los sistemas AFIS de última generación incorporan cada vez con mayor frecuencia esta lógica multimodal, apoyada en inteligencia artificial capaz de procesar en tiempo real correlaciones complejas entre modalidades distintas, en un campo que continúa desarrollándose particularmente en su capacidad de interoperar con sistemas externos de intercambio de datos entre agencias y países.

Este avance técnico plantea un principio que conviene explicitar antes de cerrar el capítulo, y que la literatura especializada en sistemas AFIS formula de manera coincidente: pese al creciente grado de automatización de estas plataformas, la decisión final sobre cualquier proceso de identificación biométrica debe seguir recayendo en un experto humano responsable, un principio que no es meramente prudencial sino que refleja exactamente la misma lección que dejó el caso Mayfield con el que se abrió este capítulo, ahora trasladada a un contexto tecnológico mucho más sofisticado. La fusión multimodal puede mejorar de manera considerable la exactitud estadística de un sistema de identificación, pero no traslada, ni debería trasladar, la responsabilidad final de esa identificación desde el perito humano hacia la máquina.

La identificación de víctimas en contextos de desastre masivo

Un escenario donde la fusión multimodal y la inteligencia artificial aplicada a la identificación humana adquieren una relevancia particularmente urgente es la identificación de víctimas en contextos de desastre masivo, conocida internacionalmente por su sigla en inglés DVI. Terremotos, accidentes de transporte masivo, incendios de gran escala y, de manera trágicamente recurrente en varios países de la región, hallazgos de fosas clandestinas vinculadas a violencia criminal organizada o a conflictos armados, generan la necesidad de identificar simultáneamente a un número elevado de personas, frecuentemente en condiciones de descomposición avanzada, fragmentación o carbonización que impiden el reconocimiento visual directo.

En estos contextos, la combinación de estimación antropológica de edad y sexo, cotejo dental automatizado y, cuando la muestra biológica lo permite, comparación genética, ofrece una vía de identificación multidisciplinaria que ningún método aislado podría sostener por sí solo con el mismo nivel de confianza. La incorporación de inteligencia artificial en cada una de estas etapas individuales, tal como se describió a lo largo de este capítulo, puede acelerar de manera considerable un proceso que, en su forma completamente manual, resulta extremadamente lento y que en escenarios de desastre masivo compite directamente con la urgencia humanitaria de devolver los restos a las familias afectadas. Al mismo tiempo, precisamente por la urgencia y la carga emocional que caracteriza a estos procesos, resulta indispensable no ceder a la tentación de acelerar la identificación sacrificando los estándares de validación que este libro viene defendiendo desde su primer capítulo, porque una identificación errónea en este contexto no solo constituye un error técnico, sino una revictimización directa de una familia que ya atraviesa una pérdida devastadora.

La región latinoamericana enfrenta, en este terreno, un desafío de escala considerable, asociado tanto a la magnitud de la desaparición forzada y la violencia homicida en varios países como a la desigual capacidad instalada de los institutos de medicina legal para procesar identificaciones masivas con los estándares internacionales recomendados por organismos como el Comité Internacional de la Cruz Roja. La incorporación gradual de herramientas de inteligencia artificial en antropología y odontología forense, siempre que se acompañe de la validación local discutida en los apartados anteriores, ofrece una oportunidad concreta de ampliar esa capacidad instalada sin necesidad de multiplicar proporcionalmente el número de especialistas humanos disponibles, una consideración de particular relevancia práctica para instituciones forenses que operan, con frecuencia, bajo restricciones presupuestarias severas.

3.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la persistencia de un lenguaje de certeza categórica en disciplinas de identificación que carecen de una base estadística equivalente a la de la genética forense, el desempeño demográfico diferencial documentado de manera sistemática en el reconocimiento facial, la escasa representación de poblaciones latinoamericanas en los conjuntos de datos utilizados para entrenar modelos de antropología y odontología forense, y el riesgo de que la automatización de tareas de identificación erosione, en lugar de fortalecer, la revisión crítica humana que el caso Mayfield demostró ser indispensable.

A estos riesgos conviene añadir uno de naturaleza más institucional que técnica: la presión, frecuentemente proveniente de la urgencia mediática o política que rodea a ciertos casos de alto perfil, de anunciar una identificación biométrica antes de haber completado el proceso de verificación humana independiente que este capítulo ha subrayado en cada una de sus secciones. La velocidad con la que un sistema de inteligencia artificial puede generar un candidato de coincidencia no debería confundirse, en ningún caso, con la velocidad con la que una identificación puede considerarse científicamente sólida y legalmente defendible. Distinguir con claridad entre ambos tiempos, el tiempo del algoritmo y el tiempo de la verificación pericial responsable, es quizás la disciplina institucional más difícil de sostener precisamente en los casos donde la presión por resultados rápidos es mayor.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir que todo resultado

de un sistema biométrico se exprese en términos probabilísticos y no de certeza categórica, requerir estudios de validación específicos sobre la población local antes de aplicar un modelo de antropología o de odontología forense desarrollado con datos de otra región del mundo, mantener siempre un proceso de revisión ciega e independiente sobre cualquier identificación generada o asistida por un algoritmo, y tratar sistemáticamente los resultados de reconocimiento facial como una pista investigativa preliminar y nunca como prueba de identificación concluyente sin corroboración adicional.

A estas prácticas conviene sumar una recomendación específica para los procesos de identificación de víctimas en contextos de desastre masivo, dado su carácter particularmente sensible desde el punto de vista humano y su relevancia para la región. Ninguna institución debería adoptar un sistema de inteligencia artificial para acelerar estos procesos sin haber establecido previamente un protocolo claro de doble verificación, que combine al menos dos líneas de evidencia independientes, por ejemplo antropológica y odontológica, o antropológica y genética, antes de comunicar una identificación a los familiares de una persona desaparecida. La velocidad que la inteligencia artificial puede aportar a este tipo de procesos es, sin duda, una ventaja humanitaria genuina, pero nunca debería obtenerse al costo de reducir el número de líneas de evidencia independientes que tradicionalmente sustentan una identificación forense responsable.

CAPÍTULO 4

Genética forense de nueva generación: secuenciación masiva, fenotipado forense por ADN y genealogía genética investigativa

4.1. Caso de apertura: el caso del Golden State Killer y el nacimiento de la genealogía genética investigativa

Entre 1974 y 1986, un agresor no identificado cometió al menos trece homicidios y más de cincuenta violaciones en distintas localidades de California, en una serie de crímenes que la prensa estadounidense bautizó como los del Golden State Killer. Durante más de tres décadas, la investigación se sostuvo sobre un perfil genético completo obtenido de la escena de los hechos, sin que ese perfil arrojara ninguna coincidencia en las bases de datos policiales convencionales, que solo permiten comparar contra personas previamente registradas por haber sido arrestadas o condenadas.

En abril de 2018, investigadores de la policía de Sacramento adoptaron una estrategia distinta: cargaron el perfil genético de la escena del crimen en una plataforma pública de genealogía genética recreativa, utilizada habitualmente por particulares para explorar su árbol genealógico, y a partir de las coincidencias parciales con parientes lejanos que sí figuraban en esa base de datos, un equipo de genealogistas reconstruyó un árbol familiar completo hasta identificar a un único individuo compatible con el perfil de la escena del crimen, que resultó ser un expolicía de setenta y dos años, Joseph James DeAngelo. La detención se produjo apenas semanas después de iniciada esa búsqueda genealógica, tras décadas de investigación tradicional sin resultado.

La resolución de este caso tuvo, además, una repercusión mediática y judicial que ningún desarrollo técnico anterior de la genética forense había alcanzado, precisamente porque combinó la resonancia emocional de un caso de más de tres décadas de antigüedad con la novedad de un método de investigación que la opinión pública general no conocía hasta entonces. Esa combinación convirtió al caso en el punto de referencia obligado para cualquier discusión posterior sobre genealogía genética investigativa, tanto en los medios de comunicación como en la literatura académica especializada, y explica por qué prácticamente toda la bibliografía citada en este capítulo remite, de una manera u otra, a este episodio como el momento fundacional de la disciplina.

Este caso, ampliamente documentado en la literatura académica y periodística, marcó el nacimiento formal de la genealogía genética investigativa como técnica forense, y desde entonces ha contribuido a esclarecer varios cientos de casos adicionales de homicidio, agresión sexual e identificación de restos humanos no identificados en Estados Unidos (Guerrini et al., 2025). Su relevancia para este capítulo es doble. Por un lado, ilustra el salto cualitativo que representa la genética forense de nueva generación respecto de la genética forense clásica basada exclusivamente en el cotejo directo de un perfil contra una base de datos de sospechosos conocidos. Por otro lado, y de manera igualmente decisiva, el caso desencadenó un intenso debate público y académico sobre los límites éticos y legales de utilizar, con fines de investigación criminal, bases de datos genéticas creadas originalmente con un propósito completamente distinto y ajeno al sistema de justicia penal, debate que atraviesa buena parte de este capítulo.

Conviene subrayar un matiz técnico que a menudo se pierde en la cobertura periodística de este tipo de casos, y que resulta decisivo para comprender correctamente el capítulo que sigue. La genealogía genética investigativa no identificó directamente al sospechoso comparando su propio perfil genético contra la base de datos, porque el sospechoso nunca había aportado su material genético a ninguna plataforma de este tipo. Lo que la técnica permitió fue identificar a parientes lejanos, hasta primos terceros o cuartos, que sí habían participado voluntariamente en esas plataformas con fines completamente ajenos a la investigación criminal, y a partir de esas coincidencias parciales, reconstruir mediante genealogía documental convencional el árbol familiar completo hasta llegar a un único individuo compatible con la evidencia genética y con el perfil demográfico del caso. Esta característica, la posibilidad de identificar a una persona a

través del consentimiento de terceros que nunca fueron consultados sobre ese uso específico, es la que distingue estructuralmente a la genealogía genética investigativa de cualquier otra disciplina descrita en este libro, y la que explica por qué su regulación resulta tan compleja.

4.2. La genética forense más allá del perfil STR

Durante más de dos décadas, la genética forense se sostuvo casi exclusivamente sobre el análisis de repeticiones cortas en tándem, conocidas por su sigla en inglés STR, un conjunto reducido de marcadores del ADN nuclear que, comparados entre dos muestras, permiten estimar con extraordinaria precisión estadística la probabilidad de que ambas provengan de la misma persona. Este método, validado exhaustivamente y respaldado por décadas de investigación poblacional, es precisamente el que los informes NAS y PCAST señalaron como la única disciplina forense que había demostrado de manera rigurosa su capacidad de asociar evidencia con una fuente específica (National Research Council, 2009; President's Council of Advisors on Science and Technology, 2016).

Las últimas dos décadas, sin embargo, han visto emerger un conjunto de desarrollos que amplían considerablemente el alcance de la genética forense más allá de ese núcleo tradicional. Este capítulo examina tres de ellos: la secuenciación masiva paralela, que multiplica de manera exponencial la cantidad y el tipo de información genética que puede extraerse de una muestra; el fenotipado forense por ADN, que permite estimar rasgos de apariencia física de una persona desconocida a partir de su perfil genético; y la genealogía genética investigativa, que amplía el universo de búsqueda de una coincidencia genética desde las bases de datos policiales tradicionales hasta las bases de datos genealógicas de uso recreativo masivo. Los tres desarrollos comparten un mismo patrón: ofrecen capacidades investigativas genuinamente nuevas, y al mismo tiempo plantean interrogantes éticos y legales que la genética forense clásica, centrada en el cotejo directo de perfiles, apenas había tenido que enfrentar.

Conviene precisar, antes de avanzar, una distinción conceptual que ordena todo este capítulo y que retoma un principio ya presentado en el segundo capítulo de este libro. Las tres tecnologías analizadas aquí no reemplazan al análisis de ADN clásico basado en STR, que sigue siendo, y probablemente seguirá siendo durante mucho tiempo, el estándar de oro para el cotejo directo entre una muestra cuestionada y un sospechoso conocido, precisamente por contar con la base estadística y de validación más sólida de toda la ciencia forense. Lo que estas tres tecnologías añaden es la capacidad de operar en escenarios donde ese cotejo directo no es posible, porque no existe ningún sospechoso conocido contra quien comparar, porque la muestra se encuentra severamente degradada, o porque se necesita información adicional, como una descripción física o un árbol de parentesco, para poder identificar en primer lugar a quién comparar. Mantener presente esta distinción entre confirmación de identidad y generación de pistas investigativas resulta indispensable para no sobrestimar el peso probatorio que corresponde a cada una de estas tres tecnologías.

4.3. Secuenciación masiva paralela: qué cambia respecto a la genética forense clásica

La secuenciación masiva paralela, también conocida como secuenciación de nueva generación, es una tecnología de alto rendimiento que permite leer simultáneamente millones de fragmentos cortos de ADN, en contraste con la secuenciación tradicional de Sanger, que procesa un fragmento por vez (Ballard et al., 2020). Aplicada a la genética forense, esta tecnología ofrece tres ventajas que ningún método anterior podía combinar simultáneamente: permite analizar cientos de marcadores

genéticos distintos en una sola corrida de laboratorio, en lugar de las veinte o treinta variantes que utiliza un kit STR convencional; ofrece información a nivel de secuencia completa de cada marcador, y no únicamente de su longitud, lo que revela variantes genéticas que los métodos anteriores no podían detectar; y muestra un desempeño considerablemente mejor sobre muestras de ADN degradado o presente en cantidades mínimas, una condición extremadamente común en la evidencia forense real recuperada de restos óseos antiguos, manchas biológicas mínimas o muestras expuestas a condiciones ambientales adversas (Ballard et al., 2020).

Una encuesta realizada entre laboratorios forenses europeos documentó que, ya en 2019, cerca de la mitad de los participantes contaba con equipamiento de secuenciación masiva paralela, y una proporción adicional planeaba adquirirlo en los siguientes dos años, lo que sugiere una trayectoria de adopción sostenida en la última media década (Gross et al., 2021, citado en la literatura especializada sobre aplicaciones de esta tecnología). Un relevamiento internacional más reciente sobre la percepción de la comunidad forense hacia estas tecnologías confirma esa tendencia de adopción creciente, aunque también documenta preocupaciones persistentes sobre el costo del equipamiento, la necesidad de personal especializado en bioinformática y la ausencia, todavía en muchos países, de un marco regulatorio específico para el uso forense de esta tecnología (Foley y Oldoni, 2023).

Esta transición tecnológica no está exenta de complejidades adicionales para la práctica forense cotidiana. El volumen de datos generado por una sola corrida de secuenciación masiva paralela supera en varios órdenes de magnitud al que produce un análisis STR convencional, lo que exige capacidad computacional de almacenamiento y procesamiento considerablemente mayor, así como personal capacitado específicamente en el análisis bioinformático de esos datos, una combinación de habilidades en genética molecular y ciencia de datos que todavía resulta escasa en el mercado laboral forense de la mayoría de los países de la región. Distintas iniciativas internacionales de cooperación técnica, incluidas las promovidas por organismos policiales multilaterales, han comenzado a ofrecer programas de capacitación específicos en esta área, reconociendo que la brecha de capital humano especializado constituye, en la práctica, un obstáculo tan relevante para la adopción responsable de esta tecnología como el costo del equipamiento en sí mismo.

Para la región iberoamericana, la incorporación de secuenciación masiva paralela plantea un desafío de infraestructura considerable, dado el costo relativamente elevado del equipamiento y de los reactivos especializados, así como la necesidad de personal capacitado en bioinformática forense, una especialización todavía escasa en la mayoría de los programas de posgrado de la región. Al mismo tiempo, precisamente por su capacidad superior para trabajar con muestras degradadas o mínimas, esta tecnología ofrece un potencial particularmente relevante para contextos como la identificación de víctimas de desaparición forzada y de fosas clandestinas, donde el material genético disponible suele encontrarse en un estado de conservación considerablemente peor que el de la evidencia forense típica de otros contextos.

Un aspecto técnico que merece explicarse en términos accesibles, dado que suele generar confusión incluso entre operadores jurídicos con formación específica en genética forense, es la diferencia entre secuenciar y genotipar. El análisis STR tradicional mide únicamente la longitud de un fragmento de ADN, es decir, cuántas veces se repite una secuencia corta determinada, sin conocer en ningún momento el contenido exacto de esa secuencia. La secuenciación masiva paralela, en cambio, lee el contenido letra por letra de cada fragmento, lo que permite detectar variantes internas dentro de un mismo marcador que dos personas distintas podrían compartir en longitud mediante el método tradicional, pero que resultan distinguibles al conocerse su secuencia exacta. Esta mayor resolución no solo mejora la capacidad de discriminación entre individuos, sino que

también permite identificar mutaciones puntuales relevantes para el análisis de parentesco en casos donde el método tradicional arrojaría resultados ambiguos, una ventaja que adquiere particular relevancia en los procesos de identificación de víctimas y de reconstrucción de vínculos familiares que se abordan más adelante en este capítulo (véase la Figura 4.1).



Figura 4.1. Diferencia entre genotipar por longitud (STR) y secuenciar (MPS). Dos alelos con el mismo número de repeticiones son indistinguibles por electroforesis capilar, pero la secuenciación revela su variante interna (isoalelos). Secuencias simplificadas con fines ilustrativos. Fuente: elaboración propia.

4.4. Genotipado probabilístico y mezclas complejas de ADN

Uno de los desafíos más persistentes de la genética forense clásica es la interpretación de mezclas de ADN, muestras que contienen material genético de dos o más personas, una situación extremadamente frecuente en evidencia como manchas de sangre mixtas, muestras de contacto en objetos manipulados por varias personas o evidencia recuperada de agresiones sexuales con múltiples agresores. La interpretación manual de estas mezclas, especialmente cuando incluyen tres o más contribuyentes en proporciones desiguales, resultó ser uno de los puntos más débiles señalados por el informe PCAST de 2016, que documentó una variabilidad considerable entre laboratorios y examinadores al momento de interpretar el mismo conjunto de datos (President's Council of Advisors on Science and Technology, 2016).

Antes de la generalización del genotipado probabilístico, la práctica forense dominante consistía en aplicar reglas de umbral relativamente simples para decidir si un determinado pico en el electroferograma correspondía a una señal genética genuina o a ruido de fondo del proceso de amplificación, y en descartar directamente como no interpretable cualquier mezcla que superara un número reducido de contribuyentes probables. Este enfoque, comprensible dado el estado de la técnica disponible en su momento, tenía como consecuencia práctica que una proporción significativa de la evidencia biológica recuperada en escenas de crimen con múltiples participantes, un escenario lejos de infrecuente en la práctica real, simplemente no podía aprovecharse como prueba, más allá del costo y el esfuerzo invertidos en su recolección y procesamiento inicial.

El software de genotipado probabilístico, mencionado ya en el primer capítulo de este libro a propósito de las observaciones que distintos actores formularon sobre el informe PCAST, aplica modelos estadísticos y de aprendizaje automático para calcular la probabilidad de distintas combinaciones de genotipos que podrían explicar un patrón de mezcla observado, y expresa su resultado como una razón de verosimilitud que compara la probabilidad de los datos bajo dos

hipótesis alternativas, por ejemplo que un sospechoso determinado sea o no sea uno de los contribuyentes de la mezcla. Esta aproximación probabilística representa un avance metodológico significativo respecto de los métodos de umbral binario utilizados anteriormente, porque permite extraer información interpretativa incluso de mezclas complejas que antes se consideraban directamente no interpretables, y porque comunica explícitamente el grado de incertidumbre asociado a cada conclusión, en lugar de presentar un resultado categórico sin matices (Coble y Bright, 2019).

El propio informe PCAST reconoció esta ventaja de manera explícita, al señalar que los programas de genotipado probabilístico constituían una mejora considerable respecto de la interpretación puramente subjetiva de mezclas complejas, aunque condicionó esa valoración positiva a que cada programa específico fuera sometido a estudios de validación independientes que confirmaran su exactitud sobre distintos escenarios de mezcla (President's Council of Advisors on Science and Technology, 2016). Esta condición resulta coherente con el marco general que este libro viene desarrollando desde su primer capítulo: ningún avance metodológico, por prometedor que sea en principio, exime a sus desarrolladores ni a los laboratorios que lo adoptan de la obligación de demostrar empíricamente su validez antes de utilizarlo como prueba en un proceso penal.

Conviene añadir que el genotipado probabilístico, pese a apoyarse en cálculos estadísticos sofisticados, no está exento del problema de la caja negra descrito en el segundo capítulo de este libro. Los modelos matemáticos subyacentes a estos programas, aunque publicados en la literatura científica, involucran un número considerable de parámetros y supuestos sobre el comportamiento del proceso de amplificación de ADN en el laboratorio, cuya combinación exacta resulta difícil de reconstruir manualmente para un perito o un abogado sin formación estadística avanzada. Por esa razón, distintas jurisdicciones han comenzado a exigir que los laboratorios que utilizan estos programas puedan producir, ante una solicitud de la defensa, no solo el resultado final expresado como razón de verosimilitud, sino también documentación suficiente sobre los parámetros y las validaciones internas del software utilizado, en línea con el principio de explicabilidad como requisito de diseño discutido en el capítulo anterior.

4.5. Fenotipado forense por ADN: prediciendo la apariencia física a partir de una muestra

El fenotipado forense por ADN, conocido por su sigla en inglés FDP, es la disciplina dedicada a predecir características visibles externas de una persona, como el color de ojos, cabello y piel, así como su ascendencia biogeográfica y, en desarrollos más recientes y todavía menos consolidados, rasgos de la morfología facial, a partir del análisis de variantes genéticas asociadas estadísticamente a esos rasgos (Terrado-Ortuño y May, 2024). A diferencia del cotejo genético clásico, que compara dos perfiles conocidos, el fenotipado forense se aplica precisamente en los casos donde no existe ningún sospechoso conocido ni ninguna coincidencia en las bases de datos disponibles, con el objetivo de generar una descripción física orientadora que reduzca el universo de personas a investigar.

El grado de exactitud de estas predicciones varía de manera considerable según el rasgo analizado. Los sistemas más consolidados, como el conjunto de marcadores conocido como HIrisPlex-S, alcanzan niveles de exactitud altos y bien validados para la predicción del color de ojos, cabello y piel, incluso en muestras degradadas o presentes en cantidades mínimas, lo que ha permitido su aplicación exitosa en casos de identificación de restos humanos en avanzado estado de descomposición. Las predicciones de rasgos más complejos, como la morfología facial completa, la forma del cabello, la presencia de pecas o el patrón de calvicie, resultan en cambio

considerablemente menos confiables, con niveles de certeza moderados a bajos que limitan seriamente su utilidad investigativa actual (Terrado-Ortuño y May, 2024).

Una precisión adicional resulta indispensable para cualquier lector no familiarizado con la genética de poblaciones: cuando este capítulo utiliza la expresión ascendencia biogeográfica, se refiere a una estimación estadística de la proporción en que el genoma de una persona comparte variantes genéticas con distintas poblaciones de referencia continentales o regionales, y no a una categoría discreta, cerrada y mutuamente excluyente como las que habitualmente se emplean en los censos poblacionales o en el lenguaje cotidiano sobre raza o etnicidad. Un mismo individuo puede presentar, de manera perfectamente coherente desde el punto de vista genético, proporciones significativas de ascendencia compartida con varias poblaciones de referencia distintas, un resultado especialmente frecuente y esperable en las poblaciones históricamente admixturadas de América Latina, y que exige del perito y del investigador una alfabetización estadística mínima para no traducir automáticamente ese resultado en una descripción racial simplificada que la propia ciencia genética no respalda.

Esta diferencia de exactitud entre rasgos simples y rasgos complejos tiene una explicación genética de fondo que conviene comprender, porque condiciona de manera directa qué puede y qué no puede esperarse razonablemente de esta tecnología en el futuro cercano. El color de ojos, y en menor medida el color de piel, están determinados en una proporción considerable por un número relativamente reducido de variantes genéticas de efecto individual grande, lo que permite construir modelos predictivos robustos a partir de un panel acotado de marcadores. La morfología facial, en cambio, es un rasgo poligénico de arquitectura genética extraordinariamente compleja, influido por cientos o miles de variantes de efecto individual mínimo que interactúan además de manera considerable con factores ambientales durante el desarrollo, lo que explica por qué, pese a los avances notables de la genómica en la última década, la predicción forense de un rostro completo a partir de ADN sigue siendo, en el estado actual del conocimiento científico, una aspiración de investigación más que una herramienta operativa madura para su uso rutinario en la práctica pericial.

Estudios recientes han ampliado el catálogo de rasgos predecibles a partir de paneles comerciales de secuenciación masiva paralela ya utilizados en la práctica forense, al identificar variantes genéticas adicionales con capacidad predictiva sobre rasgos como la obesidad o el patrón de calvicie masculina, más allá de las aplicaciones originalmente comercializadas por los fabricantes de esos kits (Elkins et al., 2023). Este tipo de extensión de uso, más allá de lo validado y comercializado originalmente, plantea una pregunta metodológica de fondo que la comunidad científica todavía discute activamente: cada nueva capacidad predictiva añadida a un panel forense requiere, en principio, su propio proceso de validación independiente, específico para el rasgo y para la población de referencia utilizada, y no puede asumirse automáticamente válida por el solo hecho de que el panel genético subyacente ya haya sido validado para otros propósitos.

El fenotipado forense por ADN plantea, además, un conjunto de tensiones éticas que lo distinguen claramente del cotejo genético tradicional. A diferencia de un perfil STR, que no revela ninguna información sobre la apariencia, la salud o el origen de una persona más allá de su identidad, un perfil de fenotipado extrae explícitamente información sensible sobre rasgos físicos y ascendencia, con el consiguiente riesgo de que esa información se utilice para reforzar sesgos preexistentes sobre determinados grupos étnicos durante una investigación, particularmente cuando la predicción de ascendencia biogeográfica se comunica de manera imprecisa como si equivaliera a una categoría racial o étnica cerrada, en lugar de una estimación probabilística sobre proporciones de ascendencia continental. La literatura especializada coincide en que la comunicación responsable de estos

resultados a los investigadores, evitando el lenguaje categórico y enfatizando siempre el margen de incertidumbre, constituye una salvaguarda ética tan importante como la validación técnica del propio método predictivo.

Esta tensión adquiere una dimensión particular en sociedades latinoamericanas, caracterizadas históricamente por un grado considerable de mestizaje entre poblaciones de ascendencia indígena, europea y africana, entre otras. Los paneles de ascendencia biogeográfica más utilizados en la práctica forense internacional fueron desarrollados y calibrados, en su gran mayoría, sobre poblaciones de referencia de Europa, Asia oriental y África subsahariana relativamente homogéneas, con una representación considerablemente menor de las poblaciones admixturadas características de la región. Esto significa que una estimación de ascendencia biogeográfica aplicada sobre una muestra latinoamericana puede arrojar resultados con un margen de incertidumbre mayor al que sugieren las cifras de validación publicadas originalmente para esos mismos paneles, una limitación que cualquier laboratorio forense de la región debería comunicar explícitamente a los investigadores y, en su caso, a las autoridades judiciales que reciban ese resultado como parte de una investigación.

4.6. Genealogía genética investigativa: bases de datos, métodos y controversias

La genealogía genética investigativa combina el perfil genético de una muestra forense, obtenido mediante un panel de cientos de miles de variantes de un solo nucleótido en lugar de los veinte o treinta marcadores STR tradicionales, con las técnicas de la genealogía documental clásica, para identificar, a través de coincidencias parciales con parientes lejanos registrados en bases de datos genealógicas de uso recreativo, a la persona de la cual proviene la muestra original. El método, popularizado a partir del caso del Golden State Killer descrito al inicio de este capítulo, ha contribuido desde entonces a resolver varios cientos de casos criminales adicionales, así como un número considerable de identificaciones de restos humanos no identificados (Guerrini et al., 2025).

El funcionamiento de esta técnica depende críticamente de la disponibilidad de bases de datos genealógicas suficientemente pobladas. Investigaciones sobre la cobertura estadística de estas bases estiman que, con aproximadamente un millón de perfiles genealógicos disponibles en una plataforma determinada, resulta posible identificar un pariente de tercer grado o más cercano para más del noventa por ciento de la población de referencia correspondiente, lo que en la práctica significa que casi cualquier persona puede ser localizada indirectamente a través de sus familiares, hayan estos aportado o no su propio material genético a la base de datos. Esta característica estructural, la capacidad de identificar a una persona a partir del material genético de terceros que nunca dieron su consentimiento específico para ese uso, constituye el núcleo del debate ético y legal que rodea a esta técnica.

El proceso técnico de una búsqueda de genealogía genética investigativa exige, además de la carga del perfil genético en la plataforma correspondiente, un trabajo de reconstrucción documental considerablemente más laborioso que el que suele imaginarse desde fuera de la disciplina. Un genealogista forense especializado debe identificar, entre decenas o cientos de coincidencias parciales reportadas por la plataforma, cuáles corresponden a parentescos genuinamente informativos y cuáles son coincidencias estadísticamente débiles o engañosas, para luego reconstruir, mediante registros civiles, obituarios, archivos históricos y árboles genealógicos publicados por los propios usuarios, la red de parentesco completa que permita converger hacia un conjunto reducido de individuos compatibles con el perfil demográfico del caso, antes de que la investigación policial tradicional retome el proceso para confirmar la identidad mediante una nueva muestra de comparación directa. Este último paso resulta crucial y a menudo se omite en la cobertura periodística de estos casos: la genealogía genética nunca constituye, por sí sola, la prueba

de identificación final, sino únicamente la vía para generar un candidato de alta probabilidad, cuya identidad debe confirmarse posteriormente mediante el cotejo genético directo y validado que describe la sección anterior de este capítulo (véase la Figura 4.2).

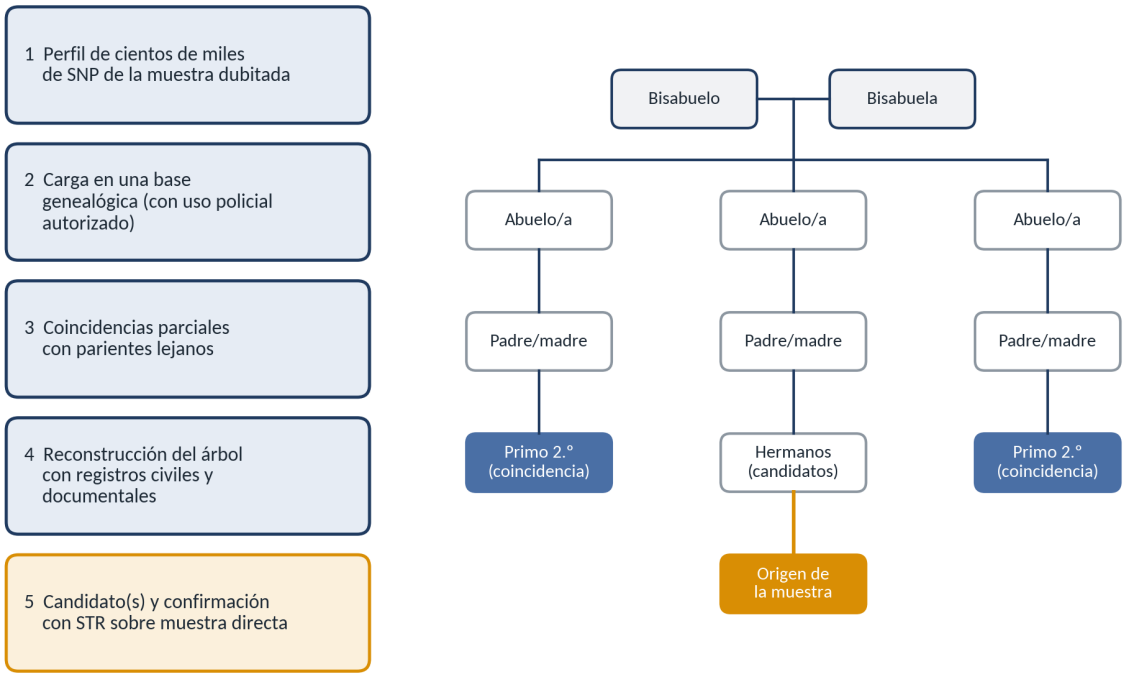


Figura 4.2. Flujo de trabajo de la genealogía genética investigativa. Las coincidencias parciales con parientes lejanos permiten reconstruir el árbol hasta un ancestro común y descender luego hacia los candidatos, cuya identidad debe confirmarse con un análisis STR convencional. Fuente: elaboración propia.

Una investigación reciente que reunió a un panel de treinta y un expertos de distintas disciplinas relevantes, mediante una metodología estructurada de consulta iterativa conocida como Delphi de política pública, identificó nueve prácticas de genealogía genética investigativa que ameritan atención regulatoria prioritaria, agrupadas en cuatro dominios: consentimiento y notificación, criterios de elegibilidad de casos, gestión de datos y privacidad, y gobernanza y rendición de cuentas (Guerrini et al., 2025). El hallazgo de mayor consenso entre los expertos consultados fue el rechazo prácticamente unánime a que las agencias de investigación accedan a bases de datos genealógicas de uso recreativo en contravención de los términos de servicio explícitamente aceptados por esas plataformas, una práctica documentada en al menos un caso conocido y que generó una controversia considerable en la comunidad de genealogía forense cuando se hizo pública en 2023.

Ese episodio, ampliamente documentado por investigaciones periodísticas especializadas, reveló que una vulnerabilidad técnica en una de las principales plataformas de genealogía genética permitió, durante varios años, que investigadores forenses accedieran a perfiles de usuarios que explícitamente habían optado por excluir sus datos de cualquier uso policial, sin que la plataforma ni los investigadores que explotaron esa vulnerabilidad lo reportaran de manera oportuna. El episodio ilustra, con particular crudeza, un problema que atraviesa buena parte de la genética forense de nueva generación descrita en este capítulo: la brecha entre lo que resulta técnicamente posible y lo que las personas cuyos datos genéticos se procesan, directa o indirectamente a través de sus familiares, efectivamente consintieron.

Un segundo hallazgo relevante del panel de expertos convocado mediante la metodología Delphi mencionada anteriormente merece destacarse aquí, porque desplaza el foco de atención desde el momento de la búsqueda genealógica propiamente dicha hacia lo que ocurre después de obtenido un resultado. Los participantes expresaron una preocupación particularmente intensa y consistente sobre la gestión posterior de los datos y las muestras biológicas generadas durante una investigación de genealogía genética, incluidos los perfiles genéticos de los familiares lejanos identificados en el proceso, que nunca fueron sospechosos de delito alguno y cuya información, sin embargo, queda incorporada de manera indefinida a los archivos de la investigación en la mayoría de los protocolos actuales. Esta preocupación por la gobernanza posterior de los datos, y no únicamente por las condiciones de acceso inicial a las bases de datos genealógicas, apunta a una laguna regulatoria que ningún país de la región ha resuelto todavía de manera integral.

4.7. Ética, privacidad y gobernanza de la genética forense de nueva generación

Las tres tecnologías descritas en este capítulo comparten un mismo desafío regulatorio de fondo: fueron desarrolladas y adoptadas por la práctica forense a una velocidad considerablemente mayor que la del marco legal específico destinado a gobernarlas. En Estados Unidos, la principal referencia regulatoria disponible hasta hace pocos años era una política interina del Departamento de Justicia, de naturaleza administrativa y no legislativa, que restringía el uso de genealogía genética investigativa a delitos violentos no resueltos y exigía el agotamiento previo de otras vías de investigación, pero que carecía del carácter vinculante y de los mecanismos de control judicial propios de una ley formal.

Distintas jurisdicciones han comenzado a avanzar hacia marcos legislativos más específicos, entre ellos una ley estatal que exige supervisión judicial previa para autorizar una búsqueda de genealogía genética investigativa, en línea con las recomendaciones formuladas por paneles de expertos convocados específicamente para estudiar esta cuestión. Este tipo de intervención legislativa, que subordina el uso de la técnica a una autorización judicial previa y no únicamente a una decisión discrecional de la agencia investigadora, ofrece un modelo de gobernanza que numerosos especialistas consideran preferible a la autorregulación puramente administrativa, precisamente porque introduce una instancia de control independiente antes de que la técnica se ponga en marcha, y no únicamente después de que sus resultados ya se hayan incorporado a una investigación en curso.

El modelo de supervisión judicial previa presenta, además, una ventaja adicional que conviene explicitar: obliga a la agencia investigadora a documentar, ante un tercero independiente, por qué las vías de investigación convencionales resultaron insuficientes y por qué el caso concreto justifica el uso de una técnica que, como se explicó anteriormente, involucra necesariamente el procesamiento de datos genéticos de terceros no sospechosos. Esta exigencia de justificación previa cumple una función análoga a la que, en otros ámbitos del proceso penal, cumple la exigencia de una orden judicial para autorizar una intervención telefónica o un allanamiento domiciliario, y traslada a la genética forense de nueva generación un estándar de control que el derecho procesal penal ya había desarrollado, con matices propios de cada tradición jurídica, para otras formas de intromisión estatal en la esfera privada de las personas.

Para la región iberoamericana, donde la genealogía genética investigativa todavía se encuentra en una etapa muy incipiente de adopción, este capítulo puede leerse como una advertencia temprana antes de que la región deba enfrentar, con años de retraso respecto de Estados Unidos, exactamente los mismos dilemas de gobernanza que allí generaron años de controversia pública, litigios y reformas legislativas apresuradas. Incorporar, desde el inicio de cualquier marco regulatorio

local sobre genética forense de nueva generación, exigencias claras de autorización judicial previa, límites explícitos sobre los delitos que justifican su uso, y protocolos de gestión de datos alineados con la legislación de protección de datos personales vigente en cada país, permitiría a las instituciones forenses de la región beneficiarse del potencial investigativo genuino de estas tecnologías sin repetir los mismos errores de gobernanza que la experiencia comparada ya ha documentado con amplitud.

Una consideración adicional, específica de la región, merece mencionarse antes de cerrar este apartado. En varios países latinoamericanos, la genética forense de nueva generación no se discute únicamente en el contexto de la investigación de delitos comunes, sino, de manera cada vez más relevante, en el contexto de la búsqueda de personas desaparecidas por motivos políticos, por violencia de género o por la actividad de organizaciones criminales. En estos escenarios, la genealogía genética investigativa y el fenotipado forense por ADN no operan como herramientas de persecución penal en sentido estricto, sino como instrumentos al servicio de un derecho humano fundamental, el derecho de las familias a conocer el destino y el paradero de sus seres queridos. Este uso humanitario de la tecnología, distinto en su finalidad última del uso propiamente criminal descrito en el resto de este capítulo, exige un marco de gobernanza propio, que equilibre de manera cuidadosa la urgencia legítima de las familias por obtener respuestas con las mismas salvaguardas de consentimiento y privacidad que este capítulo ha defendido para cualquier otro uso de estas tecnologías, y que los organismos internacionales especializados en identificación forense humanitaria han comenzado a sistematizar en protocolos específicos.

4.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la adopción de secuenciación masiva paralela y de software de genotipado probabilístico sin estudios de validación independientes específicos para cada aplicación forense concreta, la comunicación imprecisa de resultados de fenotipado y de ascendencia biogeográfica como si constituyeran categorías étnicas cerradas en lugar de estimaciones probabilísticas, el acceso a bases de datos genealógicas de uso recreativo en contravención de los términos de servicio aceptados por sus usuarios, y la ausencia, en buena parte de la región iberoamericana, de un marco legal específico que regule con claridad el uso forense de estas tres tecnologías.

Conviene añadir un riesgo de naturaleza más estructural, que atraviesa a las tres tecnologías descritas en este capítulo y que la literatura sobre gobernanza de datos genéticos identifica de manera reiterada: la asimetría de información entre, por un lado, las instituciones forenses y los proveedores comerciales de estas tecnologías, y por otro lado, el público general cuyos datos genéticos, propios o de sus familiares, terminan procesándose en el marco de una investigación criminal sin que la persona afectada tenga, en la mayoría de los casos, ningún conocimiento de ese proceso hasta mucho después de haber ocurrido. Cerrar esa asimetría de información, mediante mecanismos de transparencia institucional que informen periódicamente sobre el número y el tipo de casos en los que se emplean estas tecnologías, sin comprometer evidentemente la reserva de investigaciones en curso, constituye una condición necesaria para que la genética forense de nueva generación conserve la legitimidad social que su enorme potencial investigativo merece.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir estudios de validación independientes específicos para cada software de genotipado probabilístico antes de su adopción en

un laboratorio forense, comunicar siempre los resultados de fenotipado forense y de ascendencia biogeográfica en términos probabilísticos explícitos y nunca como categorías étnicas cerradas, limitar el uso de genealogía genética investigativa a delitos graves y no resueltos por otras vías, exigiendo autorización judicial previa siempre que el marco legal lo permita, y respetar de manera estricta los términos de servicio y las preferencias de exclusión voluntaria de las plataformas genealógicas utilizadas con fines de investigación criminal.

A estas prácticas conviene sumar dos recomendaciones adicionales de particular relevancia para la región. La primera es establecer protocolos claros de retención y destrucción de los datos genéticos de familiares identificados incidentalmente durante una búsqueda de genealogía genética investigativa, que nunca fueron sospechosos de delito alguno, en lugar de conservar esa información de manera indefinida junto con el expediente de la investigación. La segunda es que cualquier institución forense de la región que evalúe adoptar fenotipado forense por ADN encargue, antes de su implementación operativa, un estudio local que mida explícitamente el desempeño de los paneles de ascendencia biogeográfica sobre muestras representativas de la población nacional o regional correspondiente, en lugar de asumir sin verificación que las cifras de validación publicadas para poblaciones europeas o asiáticas se trasladan sin ajuste a un contexto demográfico tan distinto como el latinoamericano.



CAPÍTULO 5

La escena del crimen digital: escaneo láser 3D, fotogrametría, drones y reconstrucción virtual

5.1. Caso de apertura: cuando lo más avanzado no es necesariamente lo más convincente

Un equipo de investigadores de la Universidad de Tennessee, con el respaldo del National Institute of Justice de Estados Unidos, se propuso responder una pregunta aparentemente sencilla: ¿realmente comprenden mejor una escena del crimen quienes la observan a través de un escaneo láser tridimensional, comparado con quienes la observan mediante los métodos tradicionales de fotografía y diagramas bidimensionales? Para responderla, el equipo construyó escenas del crimen simuladas, tanto en interiores como en exteriores, y las documentó simultáneamente mediante los tres métodos más habituales de registro forense: el escaneo láser tridimensional, la fotografía convencional y el diagrama topográfico bidimensional elaborado con equipo de agrimensura (National Institute of Justice, 2022).

El estudio reclutó tanto a profesionales forenses como a personas sin formación técnica, equivalentes a los miembros de un jurado, y les formuló preguntas concretas sobre cada escena a partir de la documentación generada por cada método, para medir no solo la preferencia subjetiva de los participantes sino la exactitud objetiva de sus respuestas. El resultado desafió la intuición que suele guiar la adopción de nueva tecnología forense: si bien los participantes declararon preferir los videos e imágenes tridimensionales por resultar más inmersivos y convincentes, la exactitud efectiva de sus respuestas fue mayor cuando se apoyaron en los diagramas bidimensionales y en la fotografía convencional, mientras que el escaneo láser tridimensional, pese a su costo considerablemente mayor y a la necesidad de personal especializado para operarlo, no superó en precisión a los métodos tradicionales más económicos (National Institute of Justice, 2022).

Este hallazgo, proveniente de una institución gubernamental sin ningún interés comercial en desalentar la adopción de tecnología forense avanzada, ofrece una advertencia temprana particularmente valiosa para este capítulo. La escena del crimen digital, con sus escáneres láser, sus drones y sus reconstrucciones virtuales inmersivas, representa sin duda una transformación genuina en la capacidad de documentar y preservar evidencia espacial. Pero, como advierte el propio principio que ha organizado cada capítulo de este libro, una tecnología más sofisticada y más costosa no es automáticamente una tecnología más precisa, ni más persuasiva de manera legítima ante un tribunal, y solo la evidencia empírica específica de cada aplicación, y no la intuición sobre lo que debería funcionar mejor, puede determinar cuándo y cómo conviene adoptarla.

Conviene señalar, antes de avanzar, que este hallazgo no debe leerse como un argumento contra la innovación tecnológica en la documentación forense, sino como una ilustración particularmente clara de un principio metodológico que recorre todo este libro desde su primer capítulo: la superioridad de un método nuevo respecto de uno tradicional nunca debe asumirse por defecto, sino que debe demostrarse mediante estudios comparativos rigurosos, diseñados específicamente para medir el desempeño real en la tarea que interesa, y no únicamente la impresión subjetiva que genera en quienes lo observan. El resto de este capítulo aplicará precisamente ese estándar a cada una de las tecnologías que examina.

Conviene además situar este hallazgo en el contexto más amplio de la psicología cognitiva aplicada a la toma de decisiones judiciales. La preferencia declarada de los participantes del estudio por los formatos tridimensionales, pese a su menor exactitud efectiva, resulta consistente con investigaciones más generales sobre la relación entre fluidez de procesamiento y confianza subjetiva, según las cuales la información presentada de manera visualmente más rica y fácil de procesar tiende a generar mayor sensación de comprensión, independientemente de si esa sensación se corresponde efectivamente con una comprensión más precisa del contenido subyacente. Este fenómeno, bien documentado en la literatura sobre psicología del juicio y la decisión, ofrece una

explicación plausible del hallazgo del NIJ y anticipa, ya desde el caso de apertura de este capítulo, la discusión sobre persuasión visual que se retomará con mayor profundidad en la sección dedicada a la reconstrucción virtual.

5.2. La escena del crimen como objeto digital

La documentación de la escena del crimen constituye, desde los orígenes mismos de la criminalística descritos en el primer capítulo de este libro, una de las tareas más elementales y a la vez más determinantes de toda investigación, porque la escena es, en la inmensa mayoría de los casos, irrecuperable una vez alterada, procesada o simplemente disuelta por el paso del tiempo y de las personas que deben transitarla. Locard mismo insistió en la importancia de preservar con exactitud el estado original de una escena antes de cualquier intervención, un principio que las tecnologías descritas en este capítulo no modifican en su esencia, pero sí transforman radicalmente en su ejecución práctica.

Este capítulo examina tres tecnologías estrechamente relacionadas entre sí, el escaneo láser tridimensional, la fotogrametría y los vehículos aéreos no tripulados conocidos comúnmente como drones, que convergen en un mismo propósito: transformar la escena física, tridimensional y efímera del crimen en un objeto digital, tridimensional y persistente, que pueda examinarse, medirse y presentarse ante un tribunal mucho después de que la escena original haya dejado de existir en su forma física. El capítulo examina también la reconstrucción virtual, la etapa final de este proceso en la cual esos datos digitales se integran en representaciones interactivas destinadas tanto al análisis técnico como a la presentación ante jueces y jurados.

Antes de examinar cada tecnología en detalle, conviene precisar el vocabulario técnico básico que este capítulo empleará de manera recurrente. Una nube de puntos es el conjunto de coordenadas tridimensionales individuales capturadas por un escáner o reconstruidas por fotogrametría, cada una con su posición exacta en el espacio y, habitualmente, su color asociado. Una malla poligonal es la superficie continua que se genera al conectar esos puntos individuales mediante triángulos, produciendo una representación visualmente más completa y manipulable que la nube de puntos original. Un gemelo digital, término tomado de la ingeniería industrial y cada vez más utilizado en la literatura forense, designa a la representación tridimensional completa de una escena que aspira a replicar con fidelidad suficiente tanto la geometría como las relaciones espaciales del entorno original, de modo que las mediciones y los análisis realizados sobre esa representación digital puedan considerarse equivalentes a los que se habrían obtenido examinando la escena física original (véase la Figura 5.1).

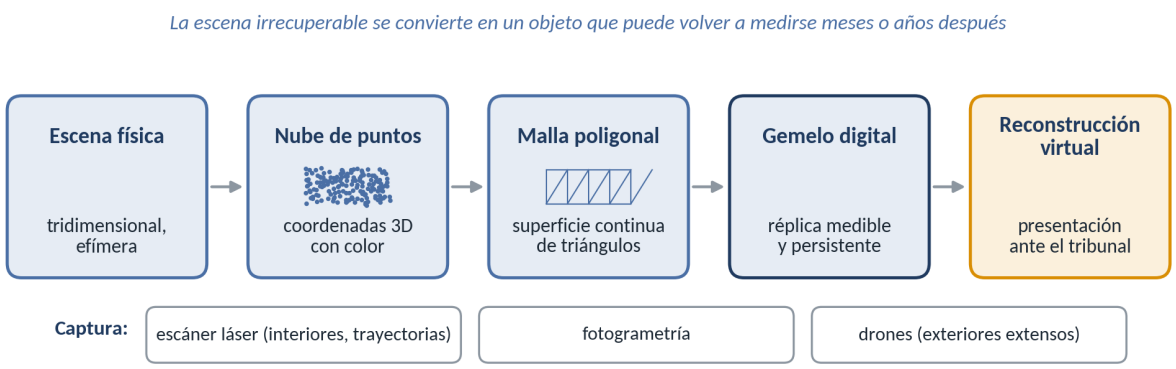


Figura 5.1. De la escena física al objeto digital: etapas y tecnologías de captura. Cada etapa conserva la geometría de la escena en un formato que puede volver a medirse después de liberado el lugar. Fuente: elaboración propia.

Conviene anticipar que las tres tecnologías de captura descritas en este capítulo no compiten necesariamente entre sí, sino que, en la práctica forense más avanzada, tienden a combinarse de manera complementaria según las características específicas de cada escena. Un accidente de tránsito en una carretera extensa puede documentarse eficientemente mediante fotogrametría aérea con drones, mientras que la trayectoria de un disparo en el interior de una vivienda puede exigir la precisión milimétrica que solo un escáner láser dedicado puede ofrecer. Aprender a seleccionar la combinación de tecnologías adecuada para cada escenario, en lugar de adoptar una única solución tecnológica de manera indiscriminada para todos los casos, es precisamente la competencia central que este capítulo intenta transmitir.

Esta transformación de la escena en un objeto digital persistente tiene, además, una implicancia procesal que merece destacarse desde el inicio de este capítulo. Bajo los métodos tradicionales de documentación, una vez procesada y liberada la escena, cualquier pregunta que surgiera posteriormente sobre un detalle no fotografiado o no medido con precisión suficiente quedaba, en la práctica, sin respuesta posible. Un modelo tridimensional completo de la escena, en cambio, permite a un perito, meses o incluso años después del hecho, realizar nuevas mediciones o verificar nuevas hipótesis sobre relaciones espaciales que nadie había considerado relevantes en el momento original de la captura, siempre que esas mediciones puedan derivarse de los datos ya capturados. Esta capacidad de interrogar retrospectivamente una escena ya inexistente constituye, quizás, el cambio más profundo que introducen estas tecnologías, y explica por qué su adopción, pese a las limitaciones documentadas en este capítulo, continúa expandiéndose de manera sostenida en la práctica forense internacional.

5.3. Escaneo láser tridimensional: fundamentos y aplicaciones

El escaneo láser terrestre funciona emitiendo pulsos de luz láser que rebotan sobre las superficies de un entorno y regresan al instrumento, que calcula la distancia exacta a cada punto de la superficie a partir del tiempo de vuelo de la señal, generando así millones de puntos individuales, conocidos colectivamente como una nube de puntos, que en conjunto reconstruyen con precisión milimétrica la geometría tridimensional completa de una escena (Balbudhe et al., 2025). Esta tecnología, empleada originalmente en topografía y arquitectura, comenzó a adoptarse en la investigación criminal a comienzos de la década de 2000, y hoy constituye una herramienta relativamente consolidada en los laboratorios forenses mejor equipados de todo el mundo (véase la Figura 5.2).



Figura 5.2. Representación de una nube de puntos de una escena urbana con cotas de dimensión. La imagen ilustra el aspecto que suele tener una nube de puntos a color, pero también un riesgo: su cota de altura (2,80 m) es incompatible con fachadas de dos plantas, que superan los seis metros. Una imagen generada puede imitar la apariencia de una medición sin tener sus propiedades métricas; solo el archivo original del escáner, con su registro de calibración, tiene valor pericial. Ilustración generada con inteligencia artificial con fines didácticos; no corresponde a evidencia real.

Entre sus aplicaciones documentadas más valiosas se encuentra el análisis de trayectorias balísticas, donde el escaneo tridimensional permite registrar con exactitud la posición y el ángulo de un orificio de entrada o salida de un proyectil, así como la geometría completa del espacio circundante, información que resulta indispensable para reconstruir posteriormente la posición relativa del tirador y de la víctima en el momento del disparo. Estudios recientes que evaluaron formalmente la incertidumbre de medición de estos ángulos de trayectoria mediante escáneres láser reportan niveles de precisión submilimétrica en condiciones controladas, aunque advierten que esa precisión puede degradarse de manera significativa según la distancia al objeto, el ángulo de incidencia del láser y las propiedades reflectantes de la superficie analizada (Greenwood et al., 2023).

Una limitación práctica documentada de manera consistente en la literatura especializada es que, pese a su extraordinaria precisión geométrica, el escaneo láser tradicional resulta considerablemente más lento y más costoso que la fotografía convencional, y exige personal específicamente capacitado tanto para la captura de los datos en el campo como para su posterior procesamiento en gabinete, dos factores que, como mostró el estudio del NIJ presentado como caso de apertura de este capítulo, no siempre se traducen en una ventaja proporcional de exactitud interpretativa para quienes finalmente deben utilizar esa documentación (National Institute of Justice, 2022). Estas limitaciones han impulsado, en los últimos años, la exploración de alternativas más ágiles y económicas, entre ellas los sensores de escaneo láser incorporados de manera nativa en teléfonos inteligentes de gama alta, que si bien ofrecen una resolución considerablemente menor que un escáner profesional dedicado,

permiten una captura casi instantánea y sin entrenamiento especializado, con una precisión que investigaciones recientes consideran suficiente para determinadas aplicaciones de documentación rápida en campo, aunque insuficiente todavía para tareas que exigen la máxima precisión métrica, como el análisis fino de trayectorias balísticas (Kottner et al., 2023).

La irrupción de estos sensores móviles de bajo costo plantea, además, una pregunta institucional interesante para las instituciones forenses de recursos limitados: en lugar de aspirar a adquirir un único escáner láser profesional de alto costo para toda una jurisdicción, con la consiguiente limitación de disponibilidad cuando ocurren varios hechos simultáneos, una estrategia alternativa consiste en equipar a un número mayor de investigadores de campo con dispositivos móviles capaces de una captura tridimensional básica, reservando el escáner profesional de mayor precisión para los casos que, por su gravedad o por la relevancia específica de una medición milimétrica, justifiquen esa inversión adicional de tiempo y de equipamiento especializado. Esta lógica de escalonamiento tecnológico, que combina herramientas de distinto costo y precisión según la relevancia probatoria del caso, reaparecerá en las secciones siguientes de este capítulo a propósito de la fotogrametría y de los drones.

Conviene cerrar esta sección señalando que la elección entre estas distintas opciones de escaneo láser no debería recaer únicamente en el investigador de campo en el momento del hecho, sino que idealmente debería estar predefinida en un protocolo institucional que clasifique, de antemano, qué tipo de escena y qué tipo de delito ameritan cada nivel de tecnología disponible. Un protocolo de esta naturaleza, elaborado con la participación de peritos, fiscales y responsables de laboratorio, permite además una asignación más eficiente de un recurso escaso, el tiempo del personal especializado en escaneo láser profesional, hacia los casos donde su precisión adicional resulta más probable que tenga peso decisivo sobre el resultado de la investigación, en lugar de dejar esa decisión librada al criterio individual y variable de cada investigador de turno.

5.4. Fotogrametría: reconstrucción tridimensional a partir de imágenes

La fotogrametría es la técnica que permite reconstruir la geometría tridimensional de un objeto o de un espacio a partir de un conjunto de fotografías bidimensionales tomadas desde distintos ángulos, mediante algoritmos que identifican puntos característicos comunes entre las distintas imágenes y triangulan su posición espacial exacta. A diferencia del escaneo láser, que requiere equipamiento especializado y costoso, la fotogrametría puede ejecutarse, en principio, con cualquier cámara digital de calidad razonable, lo que explica buena parte de su atractivo para instituciones forenses con recursos limitados.

Estudios comparativos que evaluaron de manera sistemática ambas tecnologías sobre las mismas escenas del crimen simuladas documentan que, si bien el escaneo láser tiende a ofrecer una precisión geométrica ligeramente superior en condiciones ideales, la fotogrametría alcanza niveles de exactitud estadísticamente comparables para buena parte de las aplicaciones forenses habituales, a una fracción del costo de equipamiento y con tiempos de captura considerablemente menores (Cunha et al., 2022). Esta comparación resulta particularmente relevante para instituciones forenses de la región iberoamericana que operan bajo restricciones presupuestarias significativas, para las cuales la fotogrametría puede representar una vía de acceso a la documentación tridimensional de escenas que el costo de un escáner láser profesional haría, en la práctica, inalcanzable.

La principal limitación de la fotogrametría reside en su sensibilidad a las condiciones de iluminación y a la presencia de superficies muy reflectantes, transparentes o con textura visual escasa y homogénea, condiciones bajo las cuales el algoritmo de reconstrucción carece de puntos característicos suficientes para triangular la geometría con exactitud. Esta limitación explica por

qué buena parte de la literatura especializada recomienda, cuando los recursos institucionales lo permiten, una estrategia combinada que utilice fotogrametría para la documentación general y rápida de la escena completa, complementada con escaneo láser dirigido específicamente a los puntos de mayor relevancia probatoria, como las trayectorias balísticas o los patrones de manchas de sangre, donde la precisión geométrica máxima resulta indispensable.

El procesamiento algorítmico detrás de la fotogrametría moderna, conocido técnicamente como estructura a partir del movimiento, combinado con un segundo proceso llamado estéreo de múltiples vistas, se apoya en técnicas de visión computacional directamente emparentadas con las descritas en el segundo capítulo de este libro. La calidad de la reconstrucción final depende, en gran medida, de la cantidad de fotografías capturadas, del grado de solapamiento entre ellas y de la variedad de ángulos desde los cuales se fotografió cada elemento de la escena, factores que un investigador de campo puede controlar de manera relativamente sencilla mediante un protocolo de captura estandarizado, sin necesidad de comprender en detalle la matemática subyacente al algoritmo de reconstrucción. Esta accesibilidad operativa, que contrasta con la mayor complejidad técnica del escaneo láser profesional, constituye una de las razones prácticas que explican la adopción creciente de la fotogrametría entre instituciones forenses de recursos limitados en distintas regiones del mundo.

Precisamente por esa accesibilidad operativa, la fotogrametría plantea también un riesgo de calidad inconsistente que conviene señalar con claridad. A diferencia de un escáner láser profesional, cuyo funcionamiento sigue parámetros técnicos relativamente estandarizados independientemente de quién lo opere, la calidad de una reconstrucción fotogramétrica depende en gran medida de la pericia del investigador que captura las fotografías, de su conocimiento del protocolo de solapamiento adecuado y de su capacidad de identificar, ya en el propio momento de la captura, condiciones de iluminación o de superficie que puedan comprometer la reconstrucción posterior. Esta dependencia del factor humano en la etapa de captura, mucho más pronunciada que en el escaneo láser automatizado, exige que cualquier institución que adopte la fotogrametría como método rutinario de documentación invierta de manera proporcional en la capacitación práctica y periódica de su personal de campo, y no únicamente en la adquisición del equipo fotográfico, por sí solo relativamente económico y accesible.

5.5. Drones y vehículos aéreos no tripulados en la investigación forense

Los vehículos aéreos no tripulados, popularmente conocidos como drones, han ampliado de manera considerable el repertorio de herramientas disponibles para la documentación forense, particularmente en escenarios de gran extensión donde los métodos terrestres tradicionales resultarían excesivamente lentos o directamente inviables. Una revisión exhaustiva de la literatura especializada identifica cinco áreas principales de aplicación forense de esta tecnología: la investigación de escenas del crimen propiamente dichas, la detección de fosas clandestinas, la investigación de accidentes de tránsito, el apoyo en desastres masivos y la detección de contaminación ambiental con fines periciales (Mohd Sabri et al., 2023).

En el terreno específico de la detección de evidencia, investigaciones experimentales han explorado la integración de drones equipados con cámaras de alta resolución y algoritmos de visión computacional capaces de identificar en tiempo real objetos de interés forense, como casquillos de bala o manchas de sangre, desde el aire, lo que permite a los investigadores priorizar rápidamente las áreas de un escenario extenso que ameritan una inspección terrestre más detallada, en lugar de recorrer sistemáticamente cada metro cuadrado del área acordonada (Georgiou et al., 2022). Esta capacidad resulta particularmente valiosa en escenarios exteriores de gran extensión, como los

que suelen generar los hallazgos de fosas clandestinas o los accidentes de tránsito con dispersión considerable de restos y evidencia.

El estudio experimental que documentó esta capacidad de detección aérea en tiempo real reportó, además, un hallazgo metodológico de interés práctico directo para cualquier institución que evalúe incorporar esta tecnología: la altura de vuelo del drone durante la captura influye de manera considerable en la capacidad del algoritmo de detectar correctamente objetos pequeños como casquillos de bala, con un equilibrio necesario entre volar más alto, lo que permite cubrir una superficie mayor en menos tiempo pero reduce la resolución efectiva sobre cada objeto individual, y volar más bajo, lo que mejora la detección de objetos pequeños a costa de un tiempo de cobertura considerablemente mayor sobre escenas extensas (Georgiou et al., 2022). Este tipo de parámetro operativo, aparentemente técnico y menor, ilustra un patrón que se repite en toda la tecnología descrita en este capítulo: la efectividad real de una herramienta en el campo depende de decisiones operativas específicas que exigen protocolos claros y personal capacitado para tomarlas correctamente, y no únicamente de las capacidades teóricas que el fabricante del equipo anuncia en sus especificaciones comerciales.

La fotogrametría aérea mediante drones, que combina las dos tecnologías descritas en las secciones anteriores de este capítulo, ha demostrado además un desempeño particularmente sólido para la reconstrucción tridimensional de escenas de gran extensión o de difícil acceso terrestre, como las que resultan de accidentes de tránsito en carreteras o de escenas ubicadas en terrenos de difícil circulación. Estudios recientes que evaluaron esta combinación en entornos tanto interiores como exteriores reportan niveles de precisión geométrica comparables a los de la fotogrametría terrestre convencional, con la ventaja adicional de una perspectiva aérea que permite documentar la disposición espacial completa de una escena extensa en una fracción del tiempo que exigiría un levantamiento terrestre exhaustivo (Sazaly et al., 2023).

La revisión de la literatura sobre esta tecnología identifica, sin embargo, una limitación llamativa que conviene destacar: pese a la evidencia disponible sobre su utilidad en la búsqueda y recuperación de cuerpos en contextos de desastre masivo, existe todavía escasa investigación específica sobre la integración formal de drones dentro de los protocolos internacionales de identificación de víctimas de desastre, una laguna que la propia revisión señala como una de las áreas prioritarias para el desarrollo futuro de la disciplina (Mohd Sabri et al., 2023). Esta laguna resulta especialmente relevante para la región iberoamericana, dada la magnitud de los procesos de búsqueda de personas desaparecidas y de identificación de restos en fosas clandestinas que varios países de la región enfrentan, y que ya se discutieron desde la perspectiva antropológica y odontológica en el tercer capítulo de este libro.

Más allá de su aplicación en la investigación de escenas del crimen propiamente dichas, conviene señalar que los propios drones pueden constituir, en sí mismos, objeto de investigación forense cuando participan como instrumento de un delito, ya sea mediante el transporte de sustancias ilícitas, la vigilancia no autorizada de un domicilio, o la interferencia con infraestructura crítica. Esta dimensión, conocida como informática forense de drones, exige competencias distintas a las descritas en este capítulo, más cercanas a la informática forense digital que se aborda en el capítulo siguiente de este libro, y recuerda que la relación entre esta tecnología y la criminalística es doble: los drones son, simultáneamente, una herramienta al servicio de la investigación forense y, cada vez con mayor frecuencia, un objeto que la propia investigación forense debe aprender a examinar.

El uso de drones en la investigación criminal plantea, además, consideraciones normativas específicas que distinguen a esta tecnología de las dos anteriores descritas en este capítulo. A diferencia del escaneo láser o la fotogrametría terrestre, que se ejecutan dentro del perímetro ya

acordonado de una escena bajo control policial directo, el vuelo de un drone puede sobrevolar propiedades vecinas, capturar imágenes de personas ajenas al hecho investigado, y en general generar preguntas de privacidad y de espacio aéreo regulado que la mayoría de los marcos normativos de aviación civil de la región todavía no contemplan de manera específica para el uso forense. Cualquier protocolo institucional de incorporación de drones a la investigación criminal debería, por tanto, coordinarse explícitamente con la autoridad de aviación civil correspondiente y con un marco claro de protección de datos personales para las imágenes capturadas incidentalmente durante el vuelo, más allá del objeto específico de la investigación.

5.6. Reconstrucción virtual y presentación ante tribunales

La etapa final del proceso descrito en este capítulo consiste en integrar los datos capturados mediante escaneo láser, fotogrametría o drones en un modelo tridimensional navegable, que puede explorarse posteriormente mediante software especializado, e incluso mediante dispositivos de realidad virtual que permiten a un investigador, un perito o, en determinados casos, a un juez o un jurado, recorrer virtualmente una escena que físicamente ya no existe en su estado original. La revisión sistemática sobre reconstrucción de escenas del crimen documenta que estas visualizaciones inmersivas han mostrado un valor particular en el análisis de patrones de manchas de sangre y en la confirmación de lesiones, al permitir superponer y comparar geometrías con una precisión que la observación fotográfica bidimensional no puede igualar (Balbudhe et al., 2025).

La misma revisión identifica, sin embargo, un conjunto de limitaciones que atraviesan de manera transversal toda la tecnología descrita en este capítulo: el costo considerable del equipamiento y del software especializado, las restricciones ambientales que afectan la calidad de la captura, la complejidad del procesamiento posterior de los datos, y, quizás la limitación más relevante desde la perspectiva de este libro, la ausencia de protocolos estandarizados uniformes que permitan comparar y validar de manera consistente los resultados obtenidos por distintos laboratorios y distintos operadores (Balbudhe et al., 2025). Esta ausencia de estandarización resuena directamente con la discusión sobre acreditación y estándares de calidad presentada en el primer capítulo de este libro, y confirma que la escena del crimen digital no escapa al mismo desafío de validación que atraviesa a toda la criminalística contemporánea.

Desde la perspectiva específica de la admisibilidad probatoria, la presentación de una reconstrucción virtual inmersiva ante un jurado plantea, además, un riesgo que la literatura sobre psicología jurídica ha comenzado a documentar: el llamado efecto de persuasión visual, según el cual una representación gráfica de alta fidelidad tiende a generar en quien la observa un nivel de confianza que puede superar al que objetivamente justifica la exactitud real de esa representación, precisamente la advertencia que ilustró el caso de apertura de este capítulo. Por esa razón, distintos especialistas recomiendan que toda reconstrucción virtual presentada ante un tribunal se acompañe siempre de una explicación clara y accesible de su margen de incertidumbre, de las etapas del proceso mediante el cual fue generada, y de cualquier decisión de reconstrucción que haya requerido interpretación humana antes que captura puramente objetiva de los datos, para que el impacto visual de la herramienta no sustituya, en la mente de quienes deben decidir, al análisis crítico de su validez probatoria.

Conviene distinguir, en este punto, entre dos usos posibles de una reconstrucción virtual que suelen confundirse en la práctica, pero que exigen estándares probatorios distintos. El primer uso, puramente ilustrativo, presenta la reconstrucción como una ayuda visual que acompaña el testimonio de un perito, sin pretender constituir en sí misma una prueba independiente, de manera análoga a como un diagrama dibujado en una pizarra ayuda a explicar un razonamiento sin ser,

él mismo, la prueba del hecho. El segundo uso, sustantivo, presenta la reconstrucción como una demostración de que determinada secuencia de eventos pudo o no pudo haber ocurrido dada la geometría de la escena, en cuyo caso la reconstrucción sí constituye evidencia sustantiva y debe someterse al mismo escrutinio de validación que cualquier otro método pericial descrito en este libro. Confundir estos dos usos, tratando como prueba sustantiva lo que en realidad es solamente una ayuda ilustrativa, o viceversa, constituye un riesgo probatorio que tanto los operadores judiciales como los propios peritos deben aprender a identificar con claridad.

Un desarrollo adicional que merece mencionarse en este apartado es el uso de dispositivos de realidad virtual y realidad aumentada no solo como herramienta de presentación final ante un tribunal, sino como instrumento de trabajo colaborativo durante la propia fase de investigación. Distintas iniciativas experimentales han explorado sistemas que permiten a varios investigadores, ubicados incluso en lugares físicamente distantes, explorar simultáneamente una misma reconstrucción tridimensional de una escena mediante visores de realidad aumentada, señalar puntos de interés compartidos y discutir hipótesis de reconstrucción en tiempo real sobre el propio modelo digital. Esta aplicación colaborativa, todavía en una etapa experimental de desarrollo, resulta particularmente prometedora para investigaciones complejas que requieren la participación coordinada de especialistas de distintas disciplinas forenses, como las que se describieron a propósito de la identificación de víctimas de desastre masivo en el tercer capítulo de este libro, donde antropólogos, odontólogos y genetistas forenses deben con frecuencia coordinar su trabajo sobre un mismo conjunto de hallazgos físicos dispersos en un área extensa.

La adopción de estas herramientas colaborativas, sin embargo, no está exenta de las mismas cautelas que este capítulo ha aplicado a cada tecnología examinada. La facilidad con la que un sistema de este tipo permite a múltiples personas anotar y modificar una representación compartida de la escena introduce un riesgo de trazabilidad que los protocolos tradicionales de cadena de custodia, pensados originalmente para objetos físicos individuales manipulados de manera secuencial, no siempre contemplan de manera adecuada. Cualquier institución que adopte herramientas colaborativas de este tipo debería, por tanto, exigir que el software correspondiente mantenga un registro inalterable y auditable de cada modificación introducida por cada usuario, con marca de tiempo y autoría claramente identificada, de manera que la versión final de la reconstrucción pueda reconstruirse en su historial completo de cambios si su validez llegara a cuestionarse posteriormente ante un tribunal.

5.7. Desafíos: costo, estandarización, formación y admisibilidad

El conjunto de tecnologías descritas en este capítulo comparte cuatro desafíos estructurales que conviene sistematizar antes de cerrar el capítulo. El primero es el costo, tanto del equipamiento inicial como del mantenimiento y la actualización periódica del software especializado, un desafío que afecta de manera desproporcionada a las instituciones forenses de países con menores recursos y que, como se discutió en la sección sobre fotogrametría, admite estrategias de mitigación mediante la combinación inteligente de tecnologías de distinto costo según la relevancia probatoria específica de cada elemento de la escena.

El segundo desafío es la formación especializada del personal, que debe dominar tanto la operación técnica del equipo de captura como el procesamiento posterior de los datos mediante software específico, una combinación de competencias que todavía escasea en buena parte de los programas de formación forense de la región. El tercer desafío es la ausencia de protocolos estandarizados uniformes, que dificulta comparar resultados entre laboratorios y complica la tarea de cualquier tribunal que deba evaluar la fiabilidad de una reconstrucción presentada por primera vez en

una jurisdicción determinada. El cuarto desafío, finalmente, es el riesgo de persuasión visual desproporcionada descrito en la sección anterior, que exige protocolos explícitos de comunicación de incertidumbre para cualquier reconstrucción que se presente ante un tribunal.

Frente a estos cuatro desafíos compartidos, conviene retomar el mensaje central que dejó el caso de apertura de este capítulo. Ninguna de las tecnologías descritas aquí debería adoptarse por su atractivo visual o por la sofisticación tecnológica que proyecta ante un tribunal, sino exclusivamente sobre la base de evidencia empírica específica que demuestre, para cada aplicación forense concreta, que la tecnología efectivamente mejora la exactitud de la documentación y de la interpretación de la escena, y no únicamente la impresión subjetiva de quienes la observan.

Un desafío adicional, todavía incipiente en la literatura especializada pero de creciente relevancia práctica, es la integración de inteligencia artificial en el propio proceso de captura y procesamiento de estos datos tridimensionales, mediante algoritmos capaces de segmentar automáticamente elementos relevantes de una nube de puntos, de rellenar de manera plausible zonas con datos incompletos, o de detectar automáticamente objetos de interés forense dentro de un modelo tridimensional extenso. Estos desarrollos, que combinan directamente las tecnologías descritas en este capítulo con los fundamentos de inteligencia artificial presentados en el segundo capítulo de este libro, prometen acelerar considerablemente el procesamiento de escenas complejas, pero heredan también, sin excepción, todas las exigencias de validación, transparencia y control de sesgo que ese segundo capítulo estableció como condición indispensable para cualquier herramienta algorítmica aplicada al trabajo forense.

Merece una mención particular, dentro de estos desarrollos emergentes, la técnica conocida como relleno generativo de datos faltantes, mediante la cual un algoritmo de aprendizaje profundo estima, a partir de patrones aprendidos sobre grandes volúmenes de escenas similares, cómo debería completarse una región de una nube de puntos que el escáner o la cámara no lograron capturar por completo, debido por ejemplo a una obstrucción física momentánea o a una superficie que absorbió el láser sin reflejarlo adecuadamente. Esta capacidad, sumamente útil para completar visualmente una reconstrucción con fines ilustrativos, plantea sin embargo una tensión conceptual de fondo que ningún perito debería pasar por alto: los datos generados de esta manera no son mediciones directas de la escena real, sino estimaciones estadísticas plausibles basadas en patrones aprendidos de otras escenas, y por tanto nunca deberían presentarse ante un tribunal, ni siquiera de manera implícita, como si tuvieran el mismo estatus probatorio que los datos efectivamente capturados. La distinción entre lo medido y lo estimado, ya presente de manera análoga en el capítulo anterior a propósito del fenotipado forense por ADN, reaparece aquí con una urgencia particular, precisamente porque la naturaleza visual e inmersiva de estas reconstrucciones tridimensionales facilita, más que cualquier otro tipo de evidencia forense, que esa distinción se difumine ante los ojos de quien la observa.

5.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la adopción de tecnología de documentación tridimensional por su atractivo tecnológico antes que por evidencia demostrada de su superioridad interpretativa, el riesgo de persuasión visual desproporcionada al presentar reconstrucciones inmersivas ante un jurado sin comunicar adecuadamente su margen de incertidumbre, la ausencia de protocolos estandarizados que permitan comparar resultados entre laboratorios y jurisdicciones, y la brecha de infraestructura y formación especializada que limita el acceso equitativo de las instituciones forenses de la región a estas tecnologías.

Un riesgo adicional que conviene mencionar, de naturaleza más operativa que probatoria, es la dependencia creciente de las instituciones forenses respecto de proveedores comerciales específicos de software de procesamiento tridimensional, cuyos formatos de archivo propietarios y cuyas actualizaciones periódicas pueden comprometer, con el paso de los años, la accesibilidad futura de datos capturados con versiones anteriores del mismo software. Esta forma de obsolescencia tecnológica, poco discutida en la literatura especializada pero de consecuencias prácticas considerables, plantea una pregunta que cualquier institución debería resolver antes de adoptar estas tecnologías de manera masiva: qué estrategia de preservación digital a largo plazo garantizará que un modelo tridimensional capturado hoy siga siendo accesible e interpretable dentro de diez o veinte años, un plazo perfectamente razonable considerando que muchas causas penales graves permanecen activas, o pueden reabrirse, durante períodos igualmente prolongados.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan seleccionar la tecnología de documentación según evidencia empírica específica de su desempeño para cada aplicación concreta, y no según su costo o su sofisticación aparente, combinar estratégicamente fotogrametría de menor costo con escaneo láser dirigido a los puntos de mayor relevancia probatoria, acompañar siempre cualquier reconstrucción virtual presentada ante un tribunal con una explicación clara de su margen de incertidumbre y de las decisiones de interpretación humana involucradas en su elaboración, y priorizar la inversión institucional en formación especializada de personal por sobre la sola adquisición de equipamiento avanzado.

A estas prácticas conviene sumar una recomendación final que sintetiza el argumento central de este capítulo. Antes de incorporar cualquiera de las tecnologías descritas aquí a los protocolos rutinarios de una institución forense, conviene exigir, siempre que sea posible, un estudio piloto local que compare de manera directa, como hizo el estudio del NIJ presentado al inicio de este capítulo, la tecnología candidata frente al método tradicionalmente empleado, midiendo no solo el costo y el tiempo de captura sino también la exactitud interpretativa efectiva que produce en quienes deben utilizar esa documentación, sean estos investigadores, peritos, jueces o jurados. Solo ese tipo de evidencia comparativa local, y no la reputación general de una tecnología en la literatura internacional, puede justificar de manera responsable una decisión de inversión institucional que, como se documentó a lo largo de este capítulo, suele involucrar recursos considerables y no siempre se traduce en la mejora que su atractivo tecnológico promete.



CAPÍTULO 6

Rastros materiales y algoritmos: balística, documentoscopia y análisis de trazas asistidos por IA

6.1. Caso de apertura: la balística forense responde a la crisis de validez

El primer capítulo de este libro describió cómo el informe PCAST de 2016 identificó a la identificación de armas de fuego, la disciplina que compara las marcas microscópicas que un arma deja sobre un proyectil o un casquillo disparado, como uno de los métodos forenses que carecía todavía de estudios adecuados de validez fundacional, es decir, de estudios independientes de caja negra que midieran con precisión la tasa de error de los examinadores en condiciones representativas del trabajo real (President's Council of Advisors on Science and Technology, 2016). Siete años más tarde, en mayo de 2023, un equipo de investigadores publicó en la revista Proceedings of the National Academy of Sciences uno de los estudios de mayor escala jamás realizados para responder precisamente a esa exigencia.

El estudio reclutó a doscientos veintiocho examinadores forenses de armas de fuego de distintas partes de Estados Unidos, y les pidió comparar casquillos disparados por armas conocidas bajo condiciones diseñadas para simular el trabajo de campo real. El hallazgo central resultó alentador para la disciplina: cuando se consideraban únicamente las decisiones conclusivas de identificación o eliminación, las tasas de verdadero positivo y verdadero negativo superaron el noventa y nueve por ciento, cifras que respaldan la validez fundacional del método (Guyll et al., 2023). Sin embargo, el mismo estudio reveló un matiz que el titular optimista no captura por completo: más de una quinta parte de todas las decisiones rendidas por los examinadores fueron inconclusas, es decir, ni identificaron ni eliminaron una fuente, y cuando esas decisiones inconclusas se incorporaban al cálculo general, la tasa de verdadero negativo caía de manera pronunciada hasta el sesenta y tres coma cinco por ciento, una diferencia que se explica porque los examinadores rindieron decisiones inconclusas seis veces más frecuentemente en comparaciones de fuentes distintas que en comparaciones de la misma fuente (Guyll et al., 2023).

Este hallazgo introduce el tema central de este capítulo con una precisión que difícilmente podría igualar un argumento puramente teórico. La balística forense, la documentoscopia y el análisis de trazas materiales comparten una misma naturaleza epistemológica: todas ellas dependen, en su forma tradicional, de la comparación visual de patrones por parte de un examinador humano, un proceso inherentemente subjetivo que la inteligencia artificial promete convertir en un puntaje cuantitativo, objetivo y reproducible. El estudio que abre este capítulo demuestra que esa promesa es parcialmente cierta, pero también revela que cuantificar la incertidumbre de una decisión, en lugar de eliminarla, es en muchos casos el logro más honesto y más valioso que la ciencia forense puede ofrecer.

Conviene destacar, antes de avanzar, la propia dimensión metodológica del estudio como ejemplo de buena práctica científica que este libro ha defendido en capítulos anteriores. Los autores no se limitaron a reportar una cifra global de exactitud, sino que manipularon deliberadamente la dificultad de la comparación utilizando dos modelos distintos de arma con características de marcado diferentes, precisamente para poner a prueba si la validez del método se sostenía de manera uniforme en distintas condiciones o si, por el contrario, variaba según la dificultad intrínseca de cada comparación. Este diseño experimental, mucho más exigente que limitarse a reportar una tasa de éxito favorable bajo condiciones óptimas, ejemplifica el tipo de investigación que los informes NAS y PCAST reclamaron desde 2009, y su resultado, con todos los matices que este capítulo desarrollará, constituye una de las respuestas más rigurosas que la criminalística tradicional ha producido hasta la fecha frente a esa exigencia histórica.

El propio equipo de investigación que condujo este estudio continuó defendiendo y refinando su metodología en publicaciones posteriores, respondiendo a críticas formuladas por otros estadísticos forenses sobre la manera correcta de calcular e interpretar el valor probatorio de las decisiones

inconclusas. Este intercambio académico público, con réplicas y contrarréplicas publicadas en revistas revisadas por pares, ilustra precisamente el tipo de escrutinio científico abierto que los informes NAS y PCAST reclamaron para toda la criminalística, y contrasta favorablemente con la situación descrita en el primer capítulo de este libro, en la cual buena parte de la validación tradicional de los métodos forenses se apoyaba en la aceptación profesional interna del gremio, sin el tipo de debate público y contradictorio que caracteriza a la ciencia madura.

6.2. La naturaleza de los rastros materiales: entre la comparación visual y el puntaje cuantificado

Las tres disciplinas examinadas en este capítulo, la balística, la documentoscopia y el análisis de trazas, forman parte de lo que la literatura forense denomina disciplinas de comparación de patrones, un grupo de técnicas que comparten una estructura metodológica común: un examinador humano observa dos conjuntos de marcas o de características, unas de origen conocido y otras de origen cuestionado, y emite un juicio sobre si ambas provienen o no de la misma fuente. Esta estructura común es, precisamente, la que el informe PCAST de 2016 sometió a mayor escrutinio, porque a diferencia del cotejo de perfiles de ADN, que se apoya en modelos estadísticos poblacionales rigurosos, la comparación de patrones dependió tradicionalmente del juicio experto, entrenado pero irreductiblemente subjetivo, de un examinador humano.

La incorporación de inteligencia artificial a estas disciplinas persigue un objetivo común y relativamente sencillo de formular, aunque considerablemente más difícil de ejecutar con rigor: transformar la comparación cualitativa de patrones en un puntaje numérico de similitud, que pueda someterse a los mismos estándares de validación estadística que ya se aplican al cotejo genético. Este capítulo examina hasta qué punto ese objetivo se ha alcanzado en cada una de las tres disciplinas, y qué lecciones comunes emergen del proceso.

Antes de examinar cada disciplina en detalle, conviene señalar que este capítulo continúa deliberadamente el hilo argumental iniciado en el capítulo anterior, dedicado a la escena del crimen digital. Si aquel capítulo mostraba cómo la tecnología transforma la manera de documentar el entorno físico donde ocurrió un hecho, este capítulo muestra cómo esa misma lógica de cuantificación y validación se extiende a los objetos materiales específicos que ese entorno contiene: el casquillo disparado, el documento cuestionado, la fibra transferida. En ambos casos, el hilo conductor es el mismo que ha organizado el libro completo desde su primer capítulo: ninguna tecnología, por prometedora que sea, sustituye la necesidad de una validación empírica rigurosa, específica para cada aplicación concreta y para las condiciones reales bajo las cuales esa aplicación operará en la práctica forense cotidiana.

Conviene precisar que las tres disciplinas comparten, además, un mismo punto de partida institucional que ya se discutió en el primer capítulo de este libro: a diferencia de la genética forense, que emergió directamente de la investigación académica en biología molecular, la balística, la documentoscopia y el análisis de trazas se desarrollaron principalmente dentro de laboratorios policiales orientados a resolver casos concretos, con una cultura de validación empírica considerablemente más débil que la de las ciencias biomédicas. Este origen institucional distinto ayuda a explicar por qué la incorporación de metodologías estadísticas rigurosas a estas tres disciplinas ha resultado, en términos históricos, un proceso considerablemente más tardío y más trabajoso que en la genética forense, y por qué el esfuerzo de validación descrito en este capítulo representa, en cada una de las tres áreas, un cambio cultural tan significativo como el cambio técnico que la inteligencia artificial introduce.

Vale la pena anticipar, antes de examinar cada disciplina por separado, el vocabulario estadístico

común que este capítulo empleará de manera recurrente, ya introducido de manera general en el segundo capítulo de este libro pero que conviene precisar aquí en su aplicación específica a las disciplinas de comparación de patrones. Una razón de verosimilitud expresa cuántas veces más probable resulta observar un determinado puntaje de similitud si la hipótesis de origen común es cierta, comparada con la probabilidad de observar ese mismo puntaje si la hipótesis de origen distinto es cierta. Un puntaje de similitud, obtenido mediante un algoritmo de comparación, no es todavía una razón de verosimilitud, sino un dato bruto que debe interpretarse mediante un modelo estadístico de referencia, construido a partir de miles de comparaciones previas de fuente conocida, para poder finalmente expresarse en el lenguaje probatorio que un tribunal puede comprender. Esta cadena completa, desde el puntaje bruto hasta la razón de verosimilitud final, es la que cada una de las tres disciplinas examinadas en este capítulo ha debido construir, con distintos grados de éxito, durante la última década.

6.3. Balística forense e inteligencia artificial: comparación automatizada de casquillos y proyectiles

Cuando un arma de fuego dispara, el mecanismo de percusión y el interior del cañón dejan marcas microscópicas características sobre el casquillo y el proyectil, producidas por imperfecciones individuales de fabricación y de desgaste que, según la teoría tradicional de la disciplina, resultan suficientemente distintivas como para asociar un elemento de evidencia con un arma específica. La aplicación de inteligencia artificial a esta tarea ha avanzado en dos frentes complementarios: la detección automática de las regiones de interés sobre las cuales debe realizarse la comparación, y la interpretación estadística del puntaje de similitud una vez realizada esa comparación.

En el primer frente, investigaciones recientes han desarrollado modelos de aprendizaje profundo capaces de localizar automáticamente, dentro de una imagen microscópica de un casquillo disparado, las regiones circulares correspondientes a la impresión del percutor y a la recámara, un paso preliminar que tradicionalmente exigía la intervención manual de un examinador y que introduce una fuente adicional de variabilidad entre distintos operadores (Le Bouthillier et al., 2023). En el segundo frente, sistemas de interpretación automatizada aplican modelos estadísticos de densidad, tanto paramétricos como no paramétricos, para convertir un puntaje de similitud bruto en una razón de verosimilitud que cuantifica la fuerza de la evidencia, siguiendo una lógica metodológica análoga a la descrita para el genotipado probabilístico en el capítulo anterior de este libro (Baiker-Sørensen et al., 2023).

Un hallazgo particularmente instructivo de esta línea de investigación es que el desempeño de estos sistemas automatizados no es uniforme, sino que depende de manera considerable del modelo específico de arma involucrado. Estudios que replicaron el análisis de validez fundacional con distintos modelos de pistola documentaron que ciertas marcas de fábrica producen sistemáticamente marcas microscópicas menos distintivas que otras, lo que se traduce en comparaciones más difíciles, una mayor proporción de decisiones inconclusas y una menor sensibilidad tanto para los examinadores humanos como para los sistemas algorítmicos que intentan automatizar su criterio (Guyll et al., 2023). Este hallazgo tiene una implicancia práctica directa: ningún sistema de comparación balística, humano o automatizado, debería presentarse ante un tribunal con una tasa de error única y genérica, sino que su desempeño real depende del modelo específico de arma y de las condiciones concretas de cada caso, una precisión que el propio informe PCAST ya había anticipado en 2016 al exigir estudios de validez representativos de la diversidad real de condiciones de uso.

Complementando esta línea de trabajo, investigaciones centradas en la clasificación automática

del modelo o tipo de arma a partir de las características de clase visibles en un casquillo disparado, distintas de las marcas microscópicas individuales utilizadas para la identificación propiamente dicha, han explorado arquitecturas de aprendizaje profundo capaces de distinguir entre marcas comerciales de pistola con niveles de exactitud que varían considerablemente según el fabricante, desde cifras cercanas al noventa y ocho por ciento para algunas marcas hasta niveles bastante más modestos para otras cuyas características de clase resultan menos distintivas. Esta aplicación, complementaria a la identificación individual, resulta particularmente útil en las etapas iniciales de una investigación, cuando todavía no se cuenta con un arma sospechosa específica contra la cual comparar, y el objetivo es más bien acotar el universo de armas compatibles con la evidencia recuperada en la escena.

Esta capacidad de clasificación preliminar del tipo de arma reviste una relevancia práctica particular para los contextos de investigación criminal donde el arma utilizada en un hecho nunca llega a recuperarse físicamente, un escenario lamentablemente frecuente en buena parte de la región iberoamericana, donde las armas empleadas en homicidios y otros delitos violentos suelen desecharse, ocultarse o hacerse desaparecer del circuito de investigación antes de que las autoridades puedan incautarlas. En esos casos, la única evidencia balística disponible son los propios casquillos y proyectiles recuperados en la escena, y la capacidad de un sistema algorítmico de inferir a partir de ellos el modelo probable del arma utilizada, sin necesidad de una comparación directa contra esa arma específica, puede aportar información investigativa valiosa incluso cuando la identificación individual definitiva resulte, en la práctica, inalcanzable.

6.4. El problema de las decisiones inconclusas y la interpretación algorítmica de puntajes

El hallazgo más relevante del estudio presentado como caso de apertura de este capítulo, la elevada proporción de decisiones inconclusas y su distribución asimétrica entre comparaciones de la misma fuente y de fuentes distintas, plantea un desafío metodológico que la incorporación de inteligencia artificial no resuelve automáticamente, y que la literatura especializada ha comenzado a abordar de manera directa en años recientes. Investigaciones sobre la interpretación de decisiones inconclusas demuestran que, lejos de constituir una ausencia de información, estas decisiones poseen valor probatorio propio: en el estudio de referencia, las decisiones inconclusas predijeron con una probabilidad considerable el estatus de fuente distinta, y su razón de verosimilitud asociada reforzó esa dirección (Guyll et al., 2023).

Esta observación ha llevado a distintos especialistas a proponer que los sistemas de interpretación automatizada de puntajes balísticos, en lugar de limitarse a replicar la categorización tradicional en identificación, eliminación e inconcluso, deberían diseñarse desde el inicio para producir una razón de verosimilitud continua sobre toda la gama de puntajes posibles, incluidos aquellos que hoy se clasifican como inconclusos, de manera que ningún resultado de la comparación quede sin cuantificar su aporte probatorio real (Swofford et al., 2024). Esta recomendación resulta coherente con el principio general que este libro ha defendido desde su primer capítulo: la incertidumbre no debe ocultarse tras una categoría binaria simplificada, sino comunicarse de manera explícita y cuantificada, precisamente el tipo de aporte que la inteligencia artificial, bien diseñada y validada, puede ofrecer a esta disciplina.

Una investigación complementaria advierte, sin embargo, sobre un riesgo metodológico que conviene señalar antes de continuar: el llamado efecto Hawthorne, según el cual los examinadores que saben que están siendo evaluados en el marco de un estudio de investigación tienden a modificar su comportamiento respecto del que exhibirían en su trabajo rutinario no observado, lo

que puede sesgar al alza las cifras de exactitud reportadas por los propios estudios de validación (Scurich et al., 2025). Este hallazgo introduce una capa adicional de cautela metodológica: incluso los estudios de validez fundacional mejor diseñados pueden sobrestimar el desempeño real de campo, una advertencia que refuerza, una vez más, la necesidad de mecanismos de control de calidad continuos y no únicamente de estudios de validación puntuales realizados una sola vez antes de la adopción de un método.

Esta preocupación por el efecto Hawthorne conecta de manera directa con la discusión sobre sesgo cognitivo presentada en el primer capítulo de este libro, y sugiere una vía de mitigación específica que combina ambos hallazgos de manera productiva: si los examinadores se comportan de manera distinta cuando saben que están siendo evaluados, entonces un sistema de interpretación algorítmica que opere de manera rutinaria y no anunciada, integrado al flujo normal de trabajo del laboratorio y no reservado exclusivamente para ejercicios formales de validación periódica, podría ofrecer una estimación más representativa del desempeño real de campo que la que ningún estudio puntual de validación, por riguroso que sea su diseño, puede proporcionar por sí solo. Esta posibilidad, todavía poco explorada en la literatura especializada, ilustra cómo la inteligencia artificial podría contribuir no solo a automatizar tareas específicas, sino a resolver problemas metodológicos de fondo que afectan por igual a los métodos tradicionales y a las nuevas herramientas algorítmicas que buscan complementarlos.

Un desarrollo institucional que merece mencionarse en este apartado es el papel creciente de los sistemas nacionales e internacionales de bases de datos balísticas, que permiten comparar automáticamente los casquillos y proyectiles recuperados en un hecho determinado contra un archivo histórico de miles o millones de disparos previamente registrados, con el fin de vincular hechos aparentemente aislados que en realidad involucran la misma arma. La incorporación de algoritmos de aprendizaje profundo a estos sistemas de bases de datos a gran escala ha permitido reducir de manera considerable el tiempo de procesamiento de cada consulta y ampliar el volumen de comparaciones que un laboratorio puede procesar diariamente, una capacidad operativa que resulta especialmente valiosa en contextos de violencia armada con armas de fuego de alta rotación entre distintos hechos delictivos, un escenario lamentablemente frecuente en varios países de la región iberoamericana. Sin embargo, la escala misma de estos sistemas introduce también un riesgo estadístico que conviene señalar: cuantas más comparaciones automáticas realiza un sistema de este tipo, mayor es la probabilidad acumulada de que se produzca al menos una coincidencia casual entre elementos que en realidad no comparten origen, un fenómeno estadístico conocido como el problema de las comparaciones múltiples, que exige ajustes metodológicos específicos en la interpretación de los resultados que arrojan estas plataformas de búsqueda masiva.

6.5. Documentoscopia: análisis forense de escritura y firmas asistido por inteligencia artificial

La documentoscopia forense, dedicada al análisis de escritura manuscrita, firmas y documentos cuestionados, enfrenta un desafío de validación estructuralmente similar al de la balística: durante décadas, la disciplina se apoyó en el juicio comparativo de examinadores entrenados para evaluar rasgos como la presión del trazo, la inclinación de las letras y los patrones de conexión entre caracteres, sin un marco estadístico riguroso que cuantificara la fuerza probatoria de esas observaciones. Investigaciones recientes han comenzado a desarrollar marcos estadísticos formales diseñados específicamente para asistir a los examinadores de escritura en la cuantificación objetiva de sus comparaciones, en lugar de depender exclusivamente de la experiencia acumulada y del juicio cualitativo tradicional (Crawford et al., 2023).

En paralelo a este desarrollo estadístico, la aplicación de redes neuronales profundas a la identificación de escritores ha producido sistemas capaces de extraer automáticamente características texturales y estructurales de una muestra de escritura, y de compararlas contra una base de datos de referencia con niveles de exactitud considerables en condiciones experimentales controladas (Li, 2023). Estos sistemas ofrecen, en principio, la misma promesa de objetividad y reproducibilidad que se describió para la balística: convertir la intuición visual de un examinador entrenado en un puntaje de similitud cuantificable y sometido a validación estadística formal.

La documentoscopia enfrenta, sin embargo, un desafío adicional que la distingue de la balística y que merece destacarse: a diferencia de las marcas que deja un arma de fuego, relativamente estables mientras el arma no se modifique físicamente, la escritura manuscrita de una misma persona varía de manera natural según su estado emocional, su postura, la velocidad con la que escribe y el paso del tiempo, una variabilidad intraindividual que cualquier sistema de comparación, humano o algorítmico, debe aprender a distinguir de la variabilidad interindividual que realmente permite diferenciar a un escritor de otro. Esta distinción, bien conocida por los examinadores forenses tradicionales pero técnicamente exigente de modelar matemáticamente, constituye uno de los principales desafíos abiertos para cualquier sistema de inteligencia artificial que aspire a automatizar de manera confiable esta disciplina.

Los sistemas más recientes de identificación de escritores intentan abordar este desafío mediante conjuntos de entrenamiento que incluyen múltiples muestras del mismo escritor capturadas en momentos distintos y bajo condiciones variadas, de manera que el modelo aprenda a reconocer qué aspectos de la escritura permanecen estables a través de esas variaciones naturales y cuáles son simplemente ruido circunstancial. Esta estrategia, análoga en su lógica general a los estudios de sensibilidad y especificidad discutidos en el segundo capítulo de este libro, exige sin embargo bases de datos de entrenamiento considerablemente más amplias y mejor documentadas que las disponibles para la mayoría de los estudios publicados hasta la fecha, una limitación que, como en varias otras disciplinas examinadas en este libro, restringe todavía la aplicabilidad práctica de estos sistemas en la rutina pericial cotidiana de la mayoría de los laboratorios forenses.

Un escenario adicional donde la documentoscopia asistida por inteligencia artificial ha mostrado particular utilidad es la detección de documentos falsificados mediante técnicas de imagen no destructivas, como la imagen hiperespectral, que capturan información espectral detallada de la tinta y del soporte de un documento más allá de lo visible al ojo humano, permitiendo identificar alteraciones, adiciones o sustituciones de texto que serían imperceptibles mediante el examen visual convencional. Esta capacidad resulta particularmente relevante para la documentoscopia de instrumentos financieros y contractuales, un tipo de pericia con creciente demanda en la práctica judicial de la región, donde la falsificación de firmas y la alteración posterior de documentos originalmente auténticos constituyen modalidades delictivas frecuentes que la inteligencia artificial, combinada con la instrumentación espectroscópica adecuada, puede ayudar a detectar con mayor sensibilidad que los métodos puramente visuales tradicionales.

Conviene señalar, finalmente, que la documentoscopia enfrenta hoy una presión adicional que ninguna de las otras dos disciplinas examinadas en este capítulo experimenta con la misma intensidad, y que conecta directamente con la discusión sobre inteligencia artificial generativa presentada en el segundo capítulo de este libro: la proliferación de herramientas capaces de generar firmas y escritura manuscrita sintética visualmente convincente, entrenadas sobre muestras auténticas de la escritura de una persona, introduce un escenario de falsificación completamente nuevo que la disciplina apenas comienza a estudiar de manera sistemática. A diferencia de una falsificación manual tradicional, que típicamente conserva rasgos identificables del verdadero autor

de la falsificación superpuestos a su imitación del estilo ajeno, una firma generada sintéticamente por un modelo de aprendizaje profundo entrenado específicamente para ese propósito podría carecer de esos rasgos delatores, lo que exigiría el desarrollo de técnicas de detección de contenido sintético análogas a las descritas para los deepfakes visuales en el segundo capítulo, pero adaptadas específicamente al dominio de la escritura manuscrita.

6.6. Análisis de trazas: fibras, vidrio, pintura y residuos de incendio

El análisis de trazas materiales, que incluye la comparación de fibras textiles, fragmentos de vidrio, partículas de pintura y residuos de líquidos inflamables recuperados de incendios intencionales, constituye una de las aplicaciones más directas del principio de intercambio de Locard presentado en el primer capítulo de este libro: todo contacto entre dos superficies deja, potencialmente, un rastro material transferible. La incorporación de inteligencia artificial a esta disciplina se ha concentrado especialmente en el análisis de residuos de incendios, donde la tarea consiste en determinar, a partir del perfil químico complejo obtenido mediante cromatografía de gases acoplada a espectrometría de masas, si una muestra de escombros contiene o no residuos de un líquido inflamable utilizado deliberadamente para provocar o acelerar un incendio.

Antes de examinar los resultados concretos de esta línea de investigación, conviene explicar brevemente por qué la determinación de la presencia de residuos de líquido inflamable reviste una importancia jurídica tan considerable. En la mayoría de los sistemas procesales, la diferencia entre un incendio accidental y un incendio intencional, calificado penalmente como incendio provocado o como parte de un homicidio agravado, depende en buena medida de esta determinación química, lo que convierte a esta tarea aparentemente técnica en un elemento central de la calificación jurídica del hecho investigado, y explica por qué la comunidad forense ha invertido un esfuerzo considerable en mejorar tanto la sensibilidad como la comunicación adecuada de la incertidumbre asociada a esta determinación.

Esta tarea resulta considerablemente más compleja de lo que parece a primera vista, porque la combustión y la pirólisis de los materiales de construcción y del mobiliario de una vivienda incendiada genera, de manera natural, muchos de los mismos compuestos químicos presentes en los líquidos inflamables comerciales, y porque el propio fuego suele evaporar parcialmente cualquier líquido inflamable utilizado, alterando su perfil químico característico. Un estudio que comparó directamente las opiniones de analistas humanos con las de un conjunto de modelos de aprendizaje automático, aplicados a veinte muestras de residuos de incendio analizadas de manera ciega por ambos, encontró que la combinación de la opinión del analista humano con la del sistema algorítmico produjo una estimación de la fuerza de la evidencia más robusta que cualquiera de las dos fuentes de información consideradas de manera aislada (Whitehead et al., 2023).

Este hallazgo resulta particularmente valioso para el argumento central de este libro, porque ofrece una de las demostraciones empíricas más claras, entre todas las disciplinas examinadas hasta ahora, de que la combinación explícita y metodológicamente formalizada de juicio humano y análisis algorítmico puede superar a cualquiera de los dos por separado, en lugar de tratarse de una elección excluyente entre reemplazar al experto o prescindir de la herramienta. El marco de lógica subjetiva empleado en ese estudio, que expresa tanto la opinión del analista como la del sistema algorítmico como una combinación cuantificada de creencia, incredulidad e incertidumbre, ofrece además un modelo metodológico replicable que otras disciplinas de comparación de patrones, incluidas la balística y la documentoscopia examinadas en este mismo capítulo, podrían adaptar para formalizar de manera similar la integración de sus propios sistemas algorítmicos con el juicio experto humano.

El análisis de fibras, vidrio y pintura, por su parte, se apoya crecientemente en técnicas espectroscópicas automatizadas, como la espectroscopía infrarroja por transformada de Fourier y la fluorescencia de rayos X, que generan perfiles químicos detallados de una muestra material y que pueden compararse de manera automatizada contra bases de datos de referencia. La comunidad forense ha avanzado, en los últimos años, en la estandarización de estos protocolos a través de guías técnicas específicas para cada tipo de material, un desarrollo normativo que, como se discutió en el primer capítulo de este libro a propósito de los estándares ISO, constituye un complemento indispensable de cualquier avance puramente algorítmico, porque sin protocolos estandarizados de captura y comparación, ningún sistema de inteligencia artificial puede entrenarse ni validarse de manera consistente entre distintos laboratorios.

Un desarrollo complementario, particularmente relevante para laboratorios forenses con presupuestos limitados, es la creciente disponibilidad de instrumentación espectroscópica portátil y de menor costo, cuya combinación con algoritmos de clasificación automática permite realizar análisis preliminares de trazas materiales directamente en la escena o en laboratorios regionales que carecen del equipamiento de referencia tradicionalmente concentrado en los institutos centrales de las capitales nacionales. Esta democratización tecnológica, si se acompaña de la validación adecuada frente a los instrumentos de referencia, podría contribuir a reducir la brecha de infraestructura entre laboratorios centrales y regionales que este libro ha señalado repetidamente como un desafío estructural para la criminalística de la región iberoamericana.

Conviene añadir, para completar el panorama de esta disciplina, que el análisis de trazas plantea un desafío de interpretación que le es propio y que lo distingue tanto de la balística como de la documentoscopia: incluso cuando el sistema algorítmico determina con exactitud que una fibra, un fragmento de vidrio o una partícula de pintura recuperados de un sospechoso comparten un origen químico y morfológico común con una muestra de referencia de la escena, esa coincidencia no equivale, por sí sola, a establecer que el sospechoso estuvo presente en la escena del hecho. La interpretación correcta de una coincidencia de trazas exige considerar también la frecuencia con la que ese tipo particular de material aparece de manera incidental en la población general, una pregunta empírica que estudios recientes patrocinados por agencias forenses estadounidenses han comenzado a abordar sistemáticamente, al documentar con qué frecuencia se encuentran fragmentos de vidrio y partículas de pintura en la vestimenta y el calzado de personas de la población general sin ninguna vinculación con un hecho delictivo, información de línea base indispensable para que cualquier sistema de inteligencia artificial pueda expresar correctamente, en términos de razón de verosimilitud, el verdadero peso probatorio de una coincidencia de trazas materiales.

6.7. Un hilo común: de la subjetividad al puntaje cuantificado

Las tres disciplinas examinadas en este capítulo, pese a operar sobre tipos de evidencia completamente distintos, comparten una misma trayectoria de transformación que conviene explicitar antes de cerrar el capítulo. Todas partieron de un método basado en el juicio comparativo cualitativo de un examinador entrenado, todas fueron señaladas por el informe PCAST de 2016 como disciplinas que requerían estudios adicionales de validación, y todas han respondido en los años posteriores mediante el desarrollo de sistemas que convierten esa comparación cualitativa en un puntaje cuantitativo sometido a validación estadística formal.

Esta trayectoria común permite formular una conclusión que atraviesa todo este capítulo y que se conecta directamente con el argumento central de este libro: la inteligencia artificial no ha resuelto, en ninguna de las tres disciplinas examinadas, el problema de fondo señalado por el informe PCAST, que era la ausencia de estudios de validez fundacional. Lo que ha hecho, en

los casos mejor documentados, es proporcionar la herramienta metodológica necesaria para que esos estudios de validación finalmente pudieran realizarse con el rigor estadístico que la crisis de validez de 2009 y 2016 exigió. La lección que este capítulo deja para el resto del libro es que la inteligencia artificial cumple su función más valiosa no cuando sustituye al examinador humano, sino cuando le proporciona el lenguaje estadístico necesario para que su juicio experto pueda finalmente someterse al mismo escrutinio científico que exige cualquier otra forma de prueba forense.

Este patrón compartido tiene, además, una implicancia institucional que merece subrayarse de cara a los capítulos siguientes de este libro. Las tres disciplinas examinadas aquí lograron avanzar hacia la validación estadística rigurosa no porque un único laboratorio o una única empresa desarrolladora de tecnología lo decidiera unilateralmente, sino porque comunidades científicas completas, integradas por examinadores forenses, estadísticos, informáticos y agencias de financiamiento público como el National Institute of Justice, coordinaron durante años un esfuerzo colectivo de investigación. Este modelo de desarrollo colaborativo y financiado públicamente, antes que la simple adopción de productos comerciales cerrados y no auditables, ofrece una lección de gobernanza tecnológica que la región iberoamericana debería considerar seriamente al planificar su propia incorporación de inteligencia artificial a la criminalística, un tema que los capítulos finales de este libro retomarán con mayor detalle institucional.

6.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la presentación de una tasa de error única y genérica para la comparación balística, sin considerar la variación documentada según el modelo específico de arma involucrado, el ocultamiento de la información probatoria genuina que contienen las decisiones inconclusas al tratarlas como ausencia total de evidencia, la sobrestimación del desempeño real de campo que pueden introducir los estudios de validación afectados por el efecto Hawthorne, y la insuficiente atención a la variabilidad intraindividual natural de la escritura manuscrita al entrenar sistemas de identificación de escritores.

Un riesgo adicional, de carácter más institucional, merece señalarse antes de cerrar este apartado. Las tres disciplinas examinadas en este capítulo alcanzaron sus avances de validación más significativos gracias a estudios financiados y coordinados por agencias públicas de investigación, con acceso a muestras de gran escala, a paneles de examinadores certificados dispuestos a participar en ejercicios de validación rigurosos, y a la capacidad institucional de sostener líneas de investigación durante varios años. Una institución forense que aspire simplemente a comprar e implementar un sistema comercial de inteligencia artificial para alguna de estas tres disciplinas, sin exigir la evidencia de validación equivalente a la que este capítulo ha descrito, corre el riesgo de adoptar una herramienta cuya apariencia de rigor estadístico, expresada en razones de verosimilitud y puntajes cuantificados, oculte en realidad la misma ausencia de validación fundacional que los informes NAS y PCAST identificaron en los métodos puramente subjetivos que esa herramienta pretende reemplazar.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir estudios de validez fundacional específicos para cada modelo de arma, y no una cifra genérica de exactitud aplicada indiscriminadamente a toda comparación balística, diseñar los sistemas de interpretación algorítmica para producir razones de verosimilitud continuas que cuantifiquen el valor probatorio

incluso de las decisiones tradicionalmente clasificadas como inconclusas, combinar de manera explícita y metodológicamente formalizada el juicio del analista humano con la opinión del sistema algorítmico en lugar de sustituir uno por el otro, y modelar de manera adecuada la variabilidad intraindividual natural de la escritura manuscrita antes de desplegar cualquier sistema de identificación de escritores en la práctica forense rutinaria.

A estas prácticas conviene sumar una recomendación transversal que sintetiza el argumento central de este capítulo, aplicable por igual a la balística, la documentoscopia y el análisis de trazas: cualquier laboratorio que evalúe adoptar un sistema de inteligencia artificial para alguna de estas tres disciplinas debería exigir, como condición previa, que el proveedor o el equipo de investigación responsable haya publicado un estudio que reporte no solo la exactitud global del sistema, sino su desempeño desagregado según las condiciones que la literatura de cada disciplina ya identificó como fuentes relevantes de variabilidad, el modelo de arma en el caso de la balística, el tipo de soporte o instrumento de escritura en el caso de la documentoscopia, y el grado de degradación térmica de la muestra en el caso del análisis de residuos de incendio. Una cifra de exactitud global, sin este desglose, oculta más de lo que revela sobre la verdadera utilidad forense de cualquier sistema algorítmico aplicado a estas disciplinas.



CAPÍTULO 7

El nuevo frente digital: informática forense ante deepfakes, criptoactivos, IoT y los límites del análisis predictivo

7.1. Caso de apertura: cuando los detectores de deepfakes se enfrentan al mundo real

Durante años, la investigación académica sobre detección de deepfakes reportó cifras de exactitud sumamente alentadoras, con modelos capaces de identificar contenido sintético en más del noventa y cinco por ciento de los casos evaluados. Esas cifras, sin embargo, se obtuvieron casi siempre sobre bases de datos académicas construidas en laboratorio, con deepfakes generados específicamente para la investigación y bajo condiciones controladas de iluminación, resolución y compresión. En 2025, un equipo de investigadores vinculado a la organización TrueMedia decidió poner a prueba esa promesa frente a la realidad, construyendo un nuevo conjunto de evaluación compuesto enteramente por deepfakes capturados en estado salvaje, es decir, contenido sintético que circuló realmente en redes sociales y en plataformas de verificación durante 2024, reunido a partir de ochenta y ocho sitios web distintos y cincuenta y dos idiomas diferentes (Chandra et al., 2025).

El resultado fue contundente y preocupante en igual medida. Cuando los modelos de detección de código abierto que habían reportado un desempeño casi perfecto sobre las bases de datos académicas tradicionales se evaluaron sobre este nuevo conjunto de deepfakes reales, su desempeño se desplomó: el área bajo la curva, la métrica estándar utilizada para resumir la capacidad discriminativa de un clasificador, cayó en un cincuenta por ciento para los modelos de video, un cuarenta y ocho por ciento para los de audio y un cuarenta y cinco por ciento para los de imagen, en comparación con los resultados que esos mismos modelos habían reportado sobre las bases de datos académicas convencionales (Chandra et al., 2025). Los detectores comerciales evaluados en el mismo estudio tuvieron un desempeño superior al de los modelos de código abierto, pero ni siquiera ellos lograron igualar la exactitud de un analista forense humano especializado en la detección de contenido sintético.

Este hallazgo funciona como advertencia central para todo este capítulo, que examina un conjunto de fenómenos digitales, los deepfakes, los criptoactivos, el Internet de las Cosas y el análisis predictivo, unidos por una misma característica: se trata de dominios tecnológicos que evolucionan a una velocidad considerablemente mayor que la de los estudios académicos diseñados para evaluarlos, lo que provoca que buena parte de la literatura científica disponible quede desactualizada casi en el momento mismo de su publicación. El resto de este capítulo explora, en cada uno de estos cuatro frentes, cómo la informática forense intenta mantenerse a la altura de esa velocidad de cambio, y qué lecciones metodológicas, ya anticipadas en capítulos anteriores de este libro, resultan indispensables para no confundir el entusiasmo tecnológico con la validez forense real.

Conviene añadir un dato adicional del mismo estudio que ilustra con particular precisión el problema de fondo que este capítulo desarrolla. Los propios autores documentaron que, entre 2023 y 2024, la cantidad de deepfakes identificados en esquemas de fraude se multiplicó aproximadamente por cuatro, y que durante 2023 se compartieron en redes sociales alrededor de quinientos mil deepfakes distintos, cifras que por sí solas bastarían para justificar la urgencia de este frente digital, pero que adquieren una dimensión más inquietante todavía al combinarse con evidencia de que la capacidad de las personas comunes para distinguir contenido auténtico de contenido sintético se ha reducido de manera constante a medida que mejora la calidad técnica de los generadores (Chandra et al., 2025). Esta combinación de crecimiento explosivo del contenido sintético y deterioro simultáneo de la capacidad humana de detección no asistida es, precisamente, la que convierte a la informática forense en la última línea de defensa capaz de distinguir de manera confiable lo real de lo fabricado.

7.2. El nuevo perímetro de la evidencia digital

La informática forense tradicional, dedicada al análisis de computadoras, teléfonos móviles y

redes informáticas, ya enfrentaba desafíos considerables de validación y estandarización antes de que emergieran los cuatro fenómenos examinados en este capítulo. Investigaciones sobre factores cognitivos en la informática forense demuestran que los mismos mecanismos de sesgo que el primer capítulo de este libro describió a propósito de la ciencia forense tradicional, incluido el sesgo de confirmación y la influencia de información contextual irrelevante, afectan también el trabajo de los examinadores de evidencia digital, un hallazgo que desmiente la intuición de que una disciplina tan técnica y aparentemente objetiva estaría exenta de los problemas humanos documentados en disciplinas más antiguas (Sunde y Dror, 2019).

Un desarrollo posterior de esa misma línea de investigación propuso un marco jerárquico para clasificar las distintas tareas de la informática forense según su grado de vulnerabilidad al sesgo cognitivo, de manera análoga al modelo escalonado de riesgo que el primer capítulo de este libro presentó a propósito de Dror (2023) para las disciplinas forenses tradicionales. Ese marco distingue entre tareas de bajo riesgo, como la extracción automatizada de metadatos de un archivo, donde el margen de interpretación subjetiva es mínimo, y tareas de alto riesgo, como la interpretación del significado o la intención detrás de una secuencia de comunicaciones digitales ambiguas, donde el examinador debe integrar considerable información contextual y donde, en consecuencia, el riesgo de sesgo resulta comparativamente mayor (Sunde y Dror, 2021). Esta misma lógica de clasificación por nivel de riesgo, ya aplicada en este libro a la biometría, la genética y las disciplinas de comparación de patrones, resulta igualmente útil para anticipar en qué etapas específicas de los cuatro frentes digitales examinados en este capítulo la incorporación de inteligencia artificial ofrece mayores garantías de objetividad, y en cuáles, por el contrario, el juicio humano experto sigue siendo insustituible.

Los cuatro fenómenos que este capítulo examina, deepfakes, criptoactivos, dispositivos del Internet de las Cosas y análisis predictivo, comparten una característica que los distingue de la evidencia digital tradicional examinada hasta ahora: no se trata simplemente de nuevos tipos de archivos o dispositivos que analizar con las mismas herramientas de siempre, sino de fenómenos que alteran de manera fundamental la propia naturaleza de lo que cuenta como evidencia confiable. Un video ya no es, por sí solo, prueba de que algo ocurrió; una transacción financiera ya no revela automáticamente la identidad de quien la realizó; un dispositivo doméstico conectado a internet puede convertirse en testigo silencioso de un hecho sin que nadie lo hubiera anticipado; y un algoritmo que promete predecir dónde ocurrirá el próximo delito puede, en la práctica, limitarse a reproducir los sesgos históricos de la vigilancia policial. Cada una de las secciones siguientes examina en detalle estos cuatro desafíos.

Conviene precisar, además, que estos cuatro fenómenos no operan de manera aislada, sino que con frecuencia convergen dentro de una misma investigación. Un esquema de fraude financiero puede combinar un video generado sintéticamente para engañar a una víctima, un flujo de criptoactivos diseñado para dificultar el rastreo de las ganancias ilícitas, y dispositivos del Internet de las Cosas comprometidos como parte de la infraestructura técnica del propio esquema delictivo. Esta convergencia exige que los peritos informáticos forenses de la próxima década desarrollen una competencia transversal en los cuatro frentes examinados en este capítulo, en lugar de especializarse exclusivamente en uno de ellos, una exigencia formativa que el capítulo de este libro dedicado al perito del futuro retomará con mayor detalle (véase la Figura 7.1).

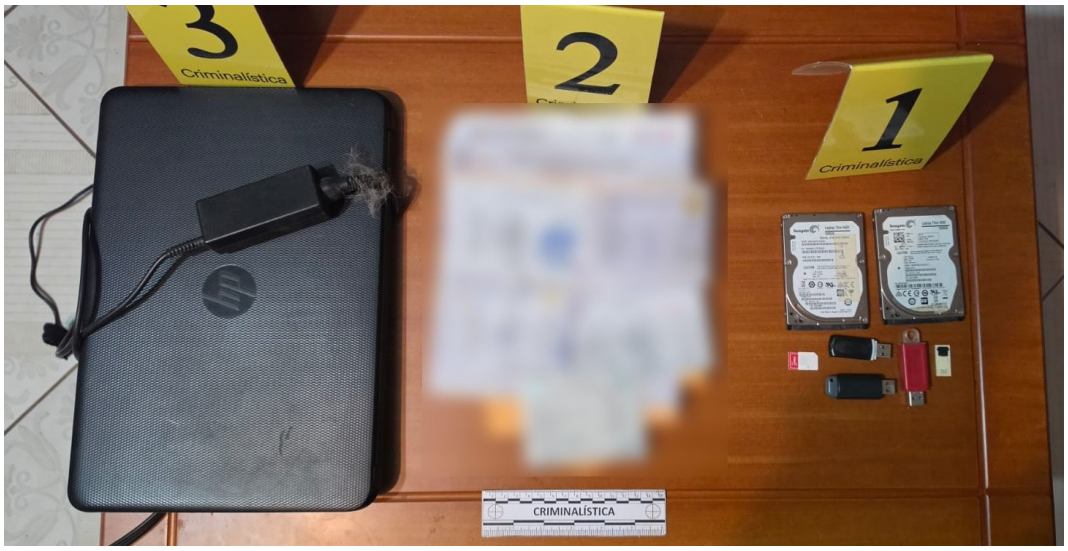


Figura 7.1. Fijación de indicios digitales con indicadores numerados y testigo métrico. En un mismo levantamiento confluyen soportes heterogéneos: equipo portátil, discos duros, memorias USB, tarjeta SIM y documentos. Los documentos se difuminaron para no exponer datos personales. Ilustración generada con inteligencia artificial con fines didácticos; no corresponde a evidencia real.

7.3. Deepfakes y contenido sintético: detección forense en la práctica real

Como se anticipó en el segundo capítulo de este libro, la inteligencia artificial generativa ha alcanzado un nivel de realismo que erosiona de manera creciente la confianza que tradicionalmente se depositaba en la evidencia audiovisual. Una revisión especializada en la detección forense de deepfakes multimodales, que combinan manipulación simultánea de imagen, audio y video, documenta que el propio ritmo de mejora de los generadores supera sistemáticamente al de los detectores, de modo que cualquier sistema de detección entrenado sobre una generación específica de deepfakes tiende a perder eficacia frente a la generación siguiente de herramientas generativas, apenas meses después de su desarrollo (Qureshi et al., 2024).

El caso de apertura de este capítulo ilustra con particular claridad una consecuencia práctica de este desfase: la brecha entre el desempeño de un detector sobre datos académicos y su desempeño sobre datos reales. Esta brecha tiene una explicación técnica relativamente sencilla de formular, aunque difícil de resolver: los conjuntos de datos académicos utilizados para entrenar la mayoría de los detectores contienen deepfakes generados con un número limitado de técnicas conocidas, mientras que el contenido sintético que circula realmente en redes sociales se produce con una diversidad de herramientas, muchas de ellas comerciales y de código cerrado, cuyas características técnicas específicas el detector nunca tuvo oportunidad de aprender durante su entrenamiento (Chandra et al., 2025).

Frente a este desafío, la práctica forense responsable no puede depender exclusivamente de un único sistema automatizado de detección, sino que debe combinar, como recomienda buena parte de la literatura especializada, múltiples técnicas complementarias: el análisis de artefactos visuales o auditivos de bajo nivel, como inconsistencias en el parpadeo o en la sincronización labial, el análisis de metadatos y de la cadena de procesamiento del archivo, y, de manera creciente, la verificación de procedencia mediante técnicas criptográficas que permiten certificar el origen de un archivo de medios desde el momento mismo de su captura. Ninguna de estas técnicas, tomada de manera aislada, ofrece garantías suficientes, y la combinación de varias, junto con el juicio de un analista forense capacitado, sigue siendo, según el propio estudio que abrió este capítulo, la

estrategia con mejor desempeño demostrado hasta la fecha.

La verificación de procedencia mencionada en el párrafo anterior merece una explicación adicional, porque representa un cambio de paradigma respecto de las técnicas de detección examinadas hasta ahora. En lugar de intentar detectar, después del hecho, si un archivo determinado fue manipulado, este enfoque incorpora, desde el momento mismo de la captura, una firma criptográfica verificable que documenta el dispositivo de origen, la fecha, la ubicación y cualquier edición posterior aplicada al archivo, de manera análoga a como una cadena de custodia física documenta cada persona que tuvo contacto con una pieza de evidencia material. Distintas coaliciones industriales han comenzado a desarrollar estándares técnicos abiertos para implementar este tipo de verificación de procedencia en cámaras y software de edición comercial, una iniciativa todavía en una etapa temprana de adopción pero que, de generalizarse, podría desplazar buena parte del problema de detección de deepfakes desde una batalla reactiva de identificación posterior hacia una estrategia preventiva de autenticación desde el origen, un cambio de enfoque que la comunidad forense sigue con considerable atención.

Un desafío adicional que la literatura especializada ha comenzado a documentar es la vulnerabilidad de los propios detectores frente a ataques adversariales diseñados específicamente para evadirlos, mediante pequeñas perturbaciones imperceptibles para el ojo humano pero suficientes para confundir al clasificador automatizado. Esta dinámica de ataque y defensa, ya mencionada en el segundo capítulo de este libro a propósito de la inteligencia artificial generativa en general, adquiere en el contexto forense una relevancia particular, porque significa que ningún sistema de detección de deepfakes debería certificarse como definitivamente confiable, sino que requiere revalidación periódica frente a las técnicas de generación más recientes disponibles en el momento de cada evaluación pericial concreta.

Desde la perspectiva de la admisibilidad probatoria, la irrupción de los deepfakes introduce además una complicación adicional que trasciende la pregunta puramente técnica de la detección: la mera existencia difundida de esta tecnología erosiona la confianza general que jueces, jurados y el público depositan en cualquier evidencia audiovisual, incluida la evidencia genuinamente auténtica, un fenómeno que la literatura especializada en derecho probatorio ha comenzado a denominar el dividendo del mentiroso, en referencia a la posibilidad de que una persona grabada de manera auténtica cometiendo un hecho delictivo pueda defenderse simplemente alegando que la grabación es un deepfake, sin necesidad de demostrarlo técnicamente, aprovechando la incertidumbre generalizada que esta tecnología ha instalado en la opinión pública. Este riesgo, distinto del riesgo técnico de detección examinado hasta aquí, exige que los sistemas de justicia desarrollen protocolos claros sobre quién tiene la carga de acreditar la autenticidad o la manipulación de una pieza de evidencia audiovisual, y con qué estándar de prueba, una discusión que el capítulo de este libro dedicado a la prueba algorítmica ante los tribunales retomará con mayor profundidad jurídica.

7.4. Criptoactivos: el mito de la anonimidad y la forense de blockchain

Durante años, una percepción extendida tanto entre delincuentes como entre el público general sostuvo que las criptomonedas ofrecían un canal de transacciones prácticamente anónimo e imposible de rastrear. La investigación forense sobre blockchain de la última década ha desmentido en gran medida esa percepción: dado que la mayoría de las cadenas de bloques públicas registran de manera permanente e inmutable cada transacción realizada, un investigador con las herramientas y la capacitación adecuadas puede reconstruir, en muchos casos, el flujo completo de fondos entre direcciones asociadas a un hecho delictivo (Agarwal et al., 2024).

La forense de blockchain se apoya en un conjunto de técnicas complementarias, entre ellas el análisis de agrupamiento de direcciones, que identifica patrones estadísticos indicativos de que múltiples direcciones distintas pertenecen en realidad a un mismo titular, el análisis de grafos de transacciones, que visualiza el flujo de fondos a través de la red para identificar puntos de entrada y salida hacia plataformas de intercambio reguladas, y la incorporación creciente de algoritmos de aprendizaje automático capaces de clasificar automáticamente direcciones sospechosas a partir de patrones de comportamiento transaccional. Una revisión sistemática reciente de la literatura sobre forense de blockchain documenta que estas técnicas han alcanzado un grado de madurez considerable para las principales criptomonedas de registro público, aunque su eficacia se reduce de manera significativa frente a las llamadas monedas de privacidad, diseñadas específicamente para ofuscar el origen y el destino de cada transacción mediante mecanismos criptográficos avanzados (Atlam y Wills, 2024) (véase la Figura 7.2).

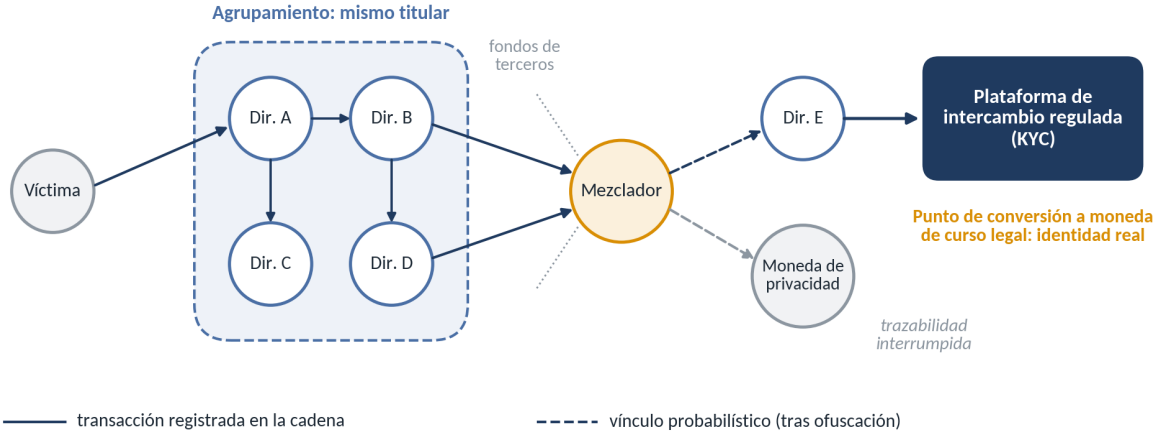


Figura 7.2. Análisis de grafos de transacciones en la forense de blockchain. El agrupamiento de direcciones y el punto de conversión en una plataforma regulada son los momentos de mayor valor investigativo; los mezcladores y las monedas de privacidad interrumpen la trazabilidad. Fuente: elaboración propia.

Esta asimetría entre criptomonedas de registro transparente y monedas de privacidad introduce una dinámica de adaptación delictiva que la literatura especializada ha comenzado a documentar empíricamente: a medida que las agencias de investigación mejoran su capacidad de rastrear criptomonedas convencionales, los actores delictivos más sofisticados migran progresivamente hacia monedas de privacidad y hacia técnicas de ofuscación adicionales, como los llamados mezcladores de criptomonedas, que combinan fondos de múltiples usuarios para dificultar la trazabilidad individual de cada transacción. Esta dinámica reproduce, en el ámbito financiero digital, la misma carrera armamentista entre generación y detección descrita en la sección anterior a propósito de los deepfakes, y confirma que ningún avance forense en este terreno puede considerarse una solución permanente, sino apenas una ventaja temporal que exige actualización continua.

Para la región iberoamericana, donde el uso de criptoactivos ha crecido de manera considerable tanto para fines legítimos de inclusión financiera como para modalidades delictivas de extorsión, fraude y lavado de activos, la incorporación de capacidades de forense de blockchain a los institutos de investigación criminal enfrenta un desafío de formación especializada considerablemente mayor que el de la informática forense tradicional, dado que exige un dominio simultáneo de criptografía, economía financiera y análisis de datos a gran escala, una combinación de competencias todavía

escasa en la mayoría de los programas de formación pericial de la región.

Un aspecto adicional que merece atención específica es el papel de las plataformas de intercambio de criptoactivos como punto de convergencia entre el mundo pseudónimo de la blockchain y la identidad real de las personas. Dado que la inmensa mayoría de las criptomonedas obtenidas de manera ilícita deben, en algún momento, convertirse en moneda de curso legal para resultar útiles a quien las obtuvo, ese momento de conversión, que típicamente ocurre a través de una plataforma de intercambio regulada sujeta a obligaciones de identificación de sus usuarios, constituye con frecuencia el punto más vulnerable de toda la cadena delictiva desde la perspectiva de la investigación forense. Las plataformas de análisis de blockchain más avanzadas han desarrollado, en consecuencia, algoritmos de aprendizaje automático especializados en identificar con alta probabilidad qué direcciones de una cadena de bloques corresponden a plataformas de intercambio conocidas, incluso cuando esas plataformas no publican esa información de manera directa, lo que permite a los investigadores concentrar sus esfuerzos precisamente en esos puntos de conversión donde la trazabilidad del flujo de fondos tiene mayores probabilidades de conducir a una identidad real.

7.5. Internet de las Cosas: la escena del crimen distribuida

El Internet de las Cosas describe la red creciente de dispositivos cotidianos conectados a internet, desde asistentes domésticos inteligentes y cámaras de seguridad hasta cerraduras electrónicas, electrodomésticos y dispositivos portátiles de monitoreo de salud, que en conjunto generan y transmiten volúmenes considerables de datos sobre la actividad cotidiana de sus usuarios. Estimaciones citadas por la literatura especializada proyectan que, hacia 2025, existirán más de treinta mil millones de conexiones de este tipo en el mundo, lo que en la práctica equivale a varios dispositivos conectados por cada persona del planeta (Friedl y Pernul, 2024).

Desde la perspectiva forense, esta proliferación transforma de manera radical la noción misma de escena del crimen descrita en el quinto capítulo de este libro: un hecho delictivo ocurrido en un domicilio equipado con dispositivos inteligentes puede haber sido registrado, sin que el investigador lo sepa inicialmente, por el historial de movimiento de una cerradura electrónica, por los registros de actividad de un asistente de voz, o por los datos de un reloj inteligente que documentó de manera involuntaria la frecuencia cardíaca de una víctima en el momento exacto de una agresión. Una revisión reciente sobre el estado de esta disciplina identifica, sin embargo, un conjunto de desafíos técnicos considerables que distinguen a la forense del Internet de las Cosas de la informática forense tradicional, entre ellos la heterogeneidad extrema de fabricantes, sistemas operativos y protocolos de comunicación, la capacidad de almacenamiento habitualmente limitada de estos dispositivos, que a menudo sobrescriben sus propios registros históricos en cuestión de días u horas, y la dispersión de los datos relevantes entre el propio dispositivo, servidores en la nube del fabricante y aplicaciones móviles asociadas, cada uno de los cuales puede estar sujeto a un régimen legal distinto de acceso y conservación (Ahmed et al., 2024).

Esta dispersión de los datos plantea un desafío de cadena de custodia particularmente complejo, porque a diferencia de un disco duro físico que el investigador puede incautar y preservar de manera directa, buena parte de la evidencia potencial de un dispositivo del Internet de las Cosas reside en servidores remotos operados por empresas privadas, frecuentemente ubicadas en jurisdicciones distintas de aquella donde ocurrió el hecho investigado, lo que exige mecanismos de cooperación legal internacional considerablemente más ágiles que los tradicionalmente disponibles para la asistencia judicial mutua entre países. Investigaciones sobre la preparación institucional para enfrentar este desafío, medida como el grado de anticipación con que una organización o

un sistema de justicia se prepara para preservar y utilizar evidencia de este tipo antes de que ocurra un hecho delictivo, documentan que la mayoría de las instituciones, incluidas las de países con sistemas forenses avanzados, todavía carecen de protocolos maduros de preparación forense específicamente diseñados para el Internet de las Cosas (Friedl y Pernul, 2024).

El concepto de preparación forense, central en esta línea de investigación, merece una explicación adicional por su relevancia práctica directa. A diferencia del enfoque tradicionalmente reactivo de la informática forense, que interviene únicamente después de ocurrido un hecho delictivo para intentar reconstruir lo sucedido a partir de la evidencia disponible, la preparación forense propone que las propias organizaciones, tanto públicas como privadas, diseñen desde el inicio sus sistemas y sus políticas de retención de datos pensando en la eventual necesidad futura de utilizar esa información con fines de investigación, por ejemplo mediante la configuración de registros de auditoría suficientemente detallados y conservados durante un período razonable, en lugar de sobrescribirlos automáticamente a los pocos días. Aplicado al Internet de las Cosas, este enfoque exigiría que los propios fabricantes de dispositivos domésticos incorporaran, desde el diseño del producto, capacidades mínimas de preservación de registros con potencial valor forense, una exigencia regulatoria que prácticamente ningún país ha incorporado todavía de manera explícita a su legislación sobre dispositivos conectados, y que representa una de las áreas de política pública más prometedoras y menos exploradas entre las examinadas en este capítulo.

La incorporación de inteligencia artificial a esta disciplina se orienta principalmente a dos tareas: la clasificación automatizada de artefactos digitales relevantes dentro del enorme volumen de datos heterogéneos que un dispositivo o un ecosistema de dispositivos puede generar, y la detección de patrones de comportamiento anómalo que podrían indicar una manipulación deliberada de los registros por parte de un sospechoso con conocimientos técnicos suficientes para intentar borrar o alterar la evidencia digital de su propio dispositivo. Ambas aplicaciones se encuentran todavía en una etapa relativamente temprana de desarrollo, y la literatura especializada coincide en que la estandarización de herramientas y protocolos de validación para esta disciplina específica constituye, por ahora, una de las principales áreas de investigación pendiente dentro de la informática forense contemporánea.

Un ejemplo concreto que ilustra tanto el potencial como los límites de esta disciplina emergente es el uso forense de dispositivos portátiles de monitoreo de salud, como relojes inteligentes y pulseras de actividad física, cuyos registros de frecuencia cardíaca, patrones de sueño y actividad física han comenzado a admitirse como evidencia en investigaciones de homicidio en distintas jurisdicciones, al permitir estimar con considerable precisión el momento aproximado de la muerte de una víctima o al contradecir la coartada temporal ofrecida por un sospechoso que también portaba uno de estos dispositivos. Sin embargo, la interpretación forense responsable de estos datos exige comprender también sus limitaciones técnicas específicas, entre ellas la posibilidad de falsos registros generados por movimientos del dispositivo sin que el usuario lo porte efectivamente puesto, y la variabilidad considerable entre distintos modelos y fabricantes en la precisión y la frecuencia de muestreo de estos sensores, consideraciones que un perito debe poder explicar con claridad ante un tribunal antes de que estos datos se conviertan en prueba central de una acusación.

7.6. Los límites reales de la inteligencia predictiva y el análisis criminal

El análisis predictivo aplicado a la investigación criminal, popularmente conocido como predictive policing, promete anticipar dónde y cuándo es más probable que ocurra un delito, con el fin de orientar de manera más eficiente el despliegue de recursos policiales limitados. Una revisión sistemática reciente, que analizó ciento sesenta y un artículos publicados sobre la efectividad de

estas herramientas basadas en grandes volúmenes de datos, ofrece un panorama considerablemente más matizado que el que suele transmitir la promoción comercial de estos sistemas, al documentar evidencia mixta e inconsistente sobre su capacidad real de reducir la criminalidad, junto con riesgos bien documentados de concentración excesiva de la vigilancia policial sobre determinados sectores de la población y de amplificación algorítmica de sesgos históricos preexistentes (Lee et al., 2024).

El problema metodológico de fondo que explica esta evidencia mixta ha sido señalado de manera reiterada por investigaciones independientes: un algoritmo de predicción criminal no aprende a partir de la totalidad de los delitos efectivamente cometidos, información que ningún sistema de justicia posee de manera completa, sino a partir de los registros policiales de arrestos y denuncias, que reflejan tanto la actividad delictiva real como las decisiones históricas de dónde y a quién la policía decidió vigilar y detener con mayor intensidad. Cuando esos registros históricos reflejan patrones de vigilancia desproporcionada sobre determinados barrios o grupos sociales, un algoritmo entrenado sobre esos mismos datos tiende a recomendar precisamente una intensificación de la vigilancia en esas mismas zonas, en un ciclo de retroalimentación que reproduce y potencialmente amplifica, en lugar de corregir, los sesgos históricos de la práctica policial que le dio origen a los datos de entrenamiento.

Una investigación periodística de gran repercusión pública, que accedió a datos operativos reales de uno de los sistemas comerciales de predicción criminal más utilizados en Estados Unidos, encontró que la tasa de acierto real de sus predicciones, medida como la proporción de predicciones que efectivamente coincidieron con un delito posterior en el lugar y horario anticipados, resultó inferior al uno por ciento en la jurisdicción analizada, una cifra dramáticamente distinta de la eficacia que la promoción comercial del sistema sugería. Este hallazgo, obtenido mediante el análisis directo de datos operativos reales y no de estudios de laboratorio controlados, ilustra con particular crudeza la brecha entre la promesa comercial de estas herramientas y su desempeño verificado en condiciones de despliegue real, un patrón que este libro ha documentado repetidamente en disciplinas muy distintas a lo largo de sus capítulos anteriores.

Esta discrepancia entre la eficacia anunciada y la eficacia real llevó, en el caso específico documentado por esa investigación periodística, a que legisladores de nivel federal solicitaran formalmente al Departamento de Justicia estadounidense que suspendiera el financiamiento público destinado a programas de este tipo, hasta contar con evidencia independiente y verificable de su efectividad. Este episodio ilustra un mecanismo de rendición de cuentas que merece destacarse como buena práctica institucional: la disponibilidad de datos operativos abiertos, susceptibles de análisis independiente por parte de investigadores y periodistas ajenos tanto al proveedor comercial como a la agencia que adquirió el sistema, resultó decisiva para que la brecha entre promesa y desempeño real saliera finalmente a la luz pública, un mecanismo de transparencia que ninguna de las otras tres tecnologías examinadas en este capítulo, deepfakes, criptoactivos e Internet de las Cosas, ha desarrollado todavía con un grado comparable de madurez institucional.

Conviene distinguir, dentro del análisis criminal asistido por inteligencia artificial, entre dos aplicaciones que la literatura especializada suele diferenciar con cuidado pero que el debate público frecuentemente confunde. El análisis predictivo basado en lugares, que estima la probabilidad de que ocurra un delito en una zona geográfica determinada durante una ventana temporal específica, es el que concentra la mayor parte de la evidencia mixta y de las críticas documentadas en esta sección. El análisis predictivo basado en personas, que intenta identificar individuos específicos con mayor probabilidad de involucrarse en actividad delictiva futura, plantea preocupaciones considerablemente más serias desde la perspectiva de los derechos fundamentales, al aproximarse

a una forma de sospecha algorítmica anticipada sobre personas que todavía no han cometido ningún delito, un terreno que distintos marcos regulatorios recientes, incluidos algunos que se examinarán en el capítulo de este libro dedicado a sesgo algorítmico y derechos humanos, han comenzado a restringir o prohibir explícitamente.

Vale la pena señalar que no toda la investigación sobre análisis predictivo arroja resultados igualmente desalentadores, y que la crítica desarrollada en esta sección no debería leerse como un rechazo categórico de cualquier uso de datos históricos para orientar la asignación de recursos policiales, una práctica que en alguna medida ha existido siempre bajo formas menos sofisticadas, como el análisis manual de mapas de criminalidad. Un estudio que replicó experimentalmente el uso de un algoritmo de predicción en la ciudad de Los Ángeles, comparando los resultados de patrullajes dirigidos algorítmicamente contra los de patrullajes tradicionales, no encontró diferencias estadísticamente significativas en la proporción de arrestos por grupo étnico entre ambas condiciones, un hallazgo que matiza, sin eliminar, la preocupación general sobre sesgo racial discutida en esta sección. Lo que esta evidencia mixta sugiere, en definitiva, es que el efecto real de estas herramientas depende de manera crítica del diseño específico del algoritmo, de la calidad y la representatividad de los datos con los que fue entrenado, y de los protocolos institucionales que rodean su implementación, y que ninguna afirmación general sobre la efectividad o el sesgo del análisis predictivo puede formularse de manera válida sin examinar esas condiciones específicas caso por caso.

7.7. Un hilo común: la velocidad del cambio tecnológico frente a la velocidad de la validación

Los cuatro fenómenos examinados en este capítulo, pese a operar en dominios técnicos completamente distintos, comparten un mismo patrón estructural que conviene explicitar antes de cerrar el capítulo: en los cuatro casos, la velocidad de desarrollo e implementación de la tecnología supera de manera sistemática a la velocidad con la que la comunidad científica puede validarla adecuadamente. Los generadores de deepfakes evolucionan más rápido que sus detectores, las técnicas de ofuscación de criptoactivos se adaptan más rápido que las herramientas de rastreo forense, la proliferación de dispositivos del Internet de las Cosas supera la capacidad institucional de desarrollar protocolos maduros de preservación de evidencia, y los sistemas de análisis predictivo se despliegan comercialmente en departamentos de policía antes de que exista evidencia científica sólida sobre su efectividad real.

Este patrón compartido confirma, con una claridad particular, la tesis central que este libro viene desarrollando desde su primer capítulo: la crisis de validez que afectó a la criminalística tradicional no desaparece con la llegada de la inteligencia artificial, sino que se reproduce, a una velocidad todavía mayor, en cada nuevo frente digital que la tecnología abre. La respuesta responsable frente a esta aceleración no es rechazar estas herramientas, cuyo potencial investigativo genuino este capítulo ha documentado en cada una de sus secciones, sino exigir, de manera todavía más sistemática que para las disciplinas forenses tradicionales, una actualización continua de los estándares de validación, precisamente porque la vida útil de cualquier hallazgo de validación en estos cuatro dominios es, por su propia naturaleza tecnológica, considerablemente más corta que en la mayoría de las disciplinas examinadas en los capítulos anteriores de este libro.

Esta observación tiene una consecuencia institucional concreta que merece formularse explícitamente. Mientras que una institución forense puede, en principio, acreditar un método de comparación balística o un protocolo de extracción de ADN y mantener esa acreditación vigente durante varios años sin necesidad de una revisión sustancial, ninguna acreditación de un sistema

de detección de deepfakes, de un protocolo de forense de blockchain o de un algoritmo de análisis predictivo debería considerarse válida de manera permanente. La propia naturaleza de estos cuatro dominios exige un modelo de acreditación continua, con ciclos de revalidación considerablemente más frecuentes que los tradicionales, un requisito que ninguno de los marcos de acreditación forense examinados en el primer capítulo de este libro, diseñados originalmente para disciplinas de evolución tecnológica mucho más lenta, contempla todavía de manera adecuada. El desarrollo de este tipo de marcos de acreditación dinámica constituye, en definitiva, una de las tareas pendientes más urgentes para la gobernanza institucional de la criminalística digital, un tema que los capítulos finales de este libro retomarán con mayor detalle normativo.

7.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la confianza excesiva en sistemas de detección de deepfakes validados exclusivamente sobre datos académicos y no sobre contenido sintético real, la subestimación de la capacidad de adaptación de los actores delictivos frente a cada nueva técnica de forense de blockchain o de detección de contenido sintético, la ausencia de protocolos maduros de preservación de evidencia distribuida en ecosistemas del Internet de las Cosas, y la adopción de sistemas de análisis predictivo criminal sin exigir evidencia independiente de su efectividad real, con el consiguiente riesgo de amplificar sesgos históricos de vigilancia desproporcionada.

Un riesgo transversal adicional, que atraviesa los cuatro frentes digitales examinados en este capítulo, es el desfase entre la velocidad de la innovación comercial y la velocidad de la formación de los operadores del sistema de justicia encargados de valorar esta evidencia. Un juez o un fiscal formado en un momento en que los deepfakes, los criptoactivos, el Internet de las Cosas o el análisis predictivo apenas existían enfrenta hoy la necesidad de evaluar críticamente prueba pericial basada en estas tecnologías, sin haber tenido necesariamente oportunidad de actualizar su formación al mismo ritmo en que la tecnología misma evoluciona. Este desfase formativo, que el capítulo de este libro dedicado al perito del futuro retomará desde la perspectiva específica de la capacitación pericial, constituye un riesgo tan estructural para la correcta valoración de esta evidencia como cualquiera de las limitaciones técnicas específicas examinadas en las secciones anteriores.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan revalidar periódicamente cualquier sistema de detección de deepfakes frente a las técnicas de generación más recientes disponibles, combinar siempre múltiples técnicas complementarias de análisis forense digital en lugar de depender de un único sistema automatizado, desarrollar protocolos institucionales de cooperación legal ágil para acceder a evidencia de dispositivos del Internet de las Cosas alojada en servidores de terceros, y exigir evidencia independiente y auditable de efectividad real, y no únicamente promoción comercial, antes de adoptar cualquier sistema de análisis predictivo criminal, prestando especial atención al riesgo de retroalimentación de sesgos históricos de vigilancia.

A estas prácticas conviene sumar una recomendación institucional final, aplicable a los cuatro frentes examinados en este capítulo: ninguna institución forense debería depender exclusivamente de herramientas comerciales cerradas para operar en estos cuatro dominios, sin mantener al mismo tiempo una capacidad propia, aunque sea modesta, de evaluación técnica independiente de esas herramientas. La opacidad comercial que caracteriza a buena parte de los productos disponibles en el mercado de detección de deepfakes, forense de blockchain, forense del Internet de las Cosas y

análisis predictivo dificulta que un laboratorio o un ministerio público pueda verificar por cuenta propia las afirmaciones de desempeño de sus proveedores, una limitación que, como demostró el caso de apertura de este capítulo, puede ocultar brechas considerables entre el desempeño anunciado y el desempeño real en condiciones de campo.



CAPÍTULO 8

Validez y derecho: la prueba algorítmica ante el tribunal

8.1. Caso de apertura: el primer video mejorado por inteligencia artificial ante un tribunal penal

En 2024, un tribunal superior del estado de Washington, Estados Unidos, se convirtió en el primero del que se tiene registro en emitir una decisión formal sobre la admisibilidad de un video mejorado mediante inteligencia artificial presentado como prueba en un proceso penal. En el caso State v. Puloka, la defensa pretendía introducir un video de vigilancia cuya claridad había sido mejorada digitalmente mediante una herramienta de inteligencia artificial, junto con el testimonio de un perito que explicara el proceso de mejora aplicado. El tribunal, aplicando el estándar Frye, que exige que una técnica científica cuente con aceptación general dentro de la comunidad especializada correspondiente antes de admitirse como prueba, excluyó tanto el video mejorado como el testimonio pericial asociado.

La resolución del tribunal fue particularmente instructiva por su razonamiento: no bastaba con que la herramienta de inteligencia artificial utilizada produjera resultados visualmente convincentes, ni con que el perito propuesto por la defensa se mostrara seguro y persuasivo al explicar su metodología. El tribunal exigió evidencia de que la técnica específica de mejora de video mediante inteligencia artificial hubiera alcanzado aceptación general dentro de la comunidad científica y forense relevante, una exigencia que la parte proponente no logró satisfacer, precisamente porque la tecnología en cuestión era demasiado reciente y su validación demasiado incipiente para haber generado ese consenso.

Resulta relevante señalar que este caso surgió de una solicitud de la propia defensa, y no de la fiscalía, un detalle que desmiente una asunción implícita frecuente en el debate público sobre inteligencia artificial y justicia penal, según la cual las herramientas algorítmicas favorecerían sistemáticamente a la acusación. En el caso Puloka, fue precisamente la defensa quien intentó introducir evidencia mejorada por inteligencia artificial con la esperanza de que esa mejora de claridad visual favoreciera a su representado, y fue el estándar de admisibilidad científica, aplicado de manera neutral por el tribunal, el que impidió que esa evidencia llegara al conocimiento del jurado sin la validación correspondiente. Este matiz confirma que un estándar de admisibilidad riguroso protege, en última instancia, a ambas partes del proceso por igual, al exigir que cualquier prueba algorítmica, provenga de donde provenga, satisfaga el mismo umbral mínimo de confiabilidad antes de poder influir en la decisión final.

Este caso, sencillo en su estructura procesal pero enormemente revelador en su contenido, sintetiza el problema central que este capítulo examina en profundidad: la inteligencia artificial no solo transforma cómo se genera y se analiza la evidencia forense, como han mostrado los capítulos anteriores de este libro, sino que exige también una transformación paralela de las categorías jurídicas mediante las cuales los tribunales deciden qué evidencia puede presentarse y con qué peso debe valorarse. Los estándares de admisibilidad probatoria desarrollados a lo largo del siglo veinte, entre ellos el propio estándar Daubert que este libro mencionó por primera vez en su primer capítulo, no anticiparon un escenario en el que la propia evidencia pudiera haber sido generada, mejorada o interpretada por un sistema algorítmico cuyo funcionamiento interno escapa, muchas veces, a la comprensión de cualquiera de las partes del proceso.

Conviene añadir un matiz procesal relevante sobre el caso Puloka que trasciende su resultado concreto. El tribunal no excluyó el video mejorado por considerar que la inteligencia artificial fuera, en abstracto, una tecnología inadecuada para el trabajo forense, una conclusión que contradiría buena parte de la evidencia presentada a lo largo de este libro sobre el potencial genuino de estas herramientas en numerosas disciplinas. Lo que el tribunal exigió, y no obtuvo, fue evidencia específica de que esa técnica particular de mejora de video hubiera sido sometida al mismo tipo

de escrutinio científico que cualquier otra técnica forense debe superar antes de presentarse ante un jurado. La lección no es que la inteligencia artificial deba excluirse de los tribunales, sino que debe ingresar a ellos por la misma puerta exigente que cualquier otra innovación científica, ni más fácilmente por su atractivo tecnológico, ni más difícilmente por la desconfianza que genera su novedad.

8.2. Qué significa prueba algorítmica ante un tribunal

A lo largo de este libro se ha utilizado la expresión prueba algorítmica para referirse a cualquier elemento probatorio que haya sido generado, procesado, interpretado o presentado con la asistencia sustancial de un sistema de inteligencia artificial, desde una identificación biométrica hasta una reconstrucción tridimensional de una escena, pasando por una razón de verosimilitud calculada mediante genotipado probabilístico o un informe de análisis de contenido sintético. Este capítulo sistematiza las exigencias jurídicas que, según la experiencia comparada examinada en las secciones siguientes, deberían aplicarse a cualquier prueba de este tipo antes de que un tribunal la considere admisible y le asigne un peso probatorio determinado.

Conviene precisar que esta sistematización no pretende agotar la totalidad de los desafíos jurídicos que la prueba algorítmica plantea, muchos de los cuales exceden el alcance específico de este libro y pertenecen propiamente al campo del derecho probatorio general. Lo que este capítulo ofrece es, más bien, un puente entre el conocimiento técnico desarrollado en los capítulos anteriores, indispensable para comprender qué es exactamente lo que un tribunal debe evaluar, y las categorías jurídicas mediante las cuales esa evaluación efectivamente se produce en la práctica judicial. Un perito, un fiscal o un defensor que haya seguido la argumentación de este libro desde su primer capítulo debería llegar a este punto con las herramientas conceptuales necesarias para formular, y para responder, las preguntas concretas que este capítulo sistematiza en las secciones siguientes.

Conviene distinguir, desde el inicio, entre dos preguntas jurídicas que con frecuencia se confunden en la práctica: la admisibilidad, que determina si una prueba puede siquiera presentarse y considerarse por el tribunal, y la valoración, que determina, una vez admitida, qué peso o qué grado de convicción esa prueba debe generar en la decisión final. Un sistema de inteligencia artificial puede superar el umbral de admisibilidad, al demostrar un mínimo de validez científica y de confiabilidad metodológica, y aun así merecer una valoración cautelosa por parte del tribunal, si su margen de error resulta considerable o si su funcionamiento interno no puede explicarse con suficiente claridad. Este capítulo aborda ambas dimensiones, aunque concentra su atención principal en la primera, por ser la que determina si el debate sobre la segunda llega siquiera a producirse.

Resulta útil, además, situar la prueba algorítmica dentro de una tipología más amplia que distinga tres roles posibles que la inteligencia artificial puede desempeñar dentro de un mismo proceso penal. En el primer rol, instrumental, el sistema algorítmico funciona como una herramienta de apoyo que un perito humano utiliza para llegar a una conclusión que sigue siendo, en última instancia, suya, como ocurre con buena parte de los sistemas de detección automática de regiones de interés examinados en capítulos anteriores. En el segundo rol, sustitutivo, el sistema algorítmico genera directamente una conclusión que se presenta ante el tribunal con una intervención humana mínima, como podría ocurrir con un resultado automatizado de un sistema de reconocimiento facial no revisado por un examinador independiente. En el tercer rol, generativo, el propio sistema produce contenido nuevo, como una imagen mejorada o una reconstrucción sintética, que se convierte en objeto directo de la prueba, como ocurrió en el caso Puloka que abrió este capítulo. Cada uno de estos tres roles exige un nivel distinto de escrutinio judicial, con el tercero requiriendo, en general,

el estándar más exigente, precisamente porque en él el sistema algorítmico no se limita a asistir un juicio humano, sino que crea directamente el objeto sobre el cual ese juicio deberá formularse.

8.3. Estándares de validación: la serie ISO 21043 y la acreditación de herramientas de inteligencia artificial

Como se mencionó en el primer capítulo de este libro, la comunidad forense internacional ha desarrollado, a través de la Organización Internacional de Normalización, un conjunto de estándares específicamente diseñados para la ciencia forense, organizados bajo la serie ISO 21043 y estructurados en cinco partes que siguen el proceso forense completo, desde la escena hasta la sala de audiencias: vocabulario común, reconocimiento y recolección de elementos, análisis, interpretación y elaboración de informes. Esta estructura, publicada de manera escalonada entre 2018 y 2025, ofrece un marco de referencia que, aunque no fue diseñado originalmente pensando en la inteligencia artificial, resulta directamente aplicable a la validación de cualquier herramienta algorítmica incorporada a alguna de esas cinco etapas del proceso forense.

Merece destacarse, además, que la propia parte de la serie dedicada al vocabulario, aparentemente la más modesta de las cinco, cumple una función poco visible pero decisiva para el argumento de este capítulo: al fijar definiciones comunes y estandarizadas para conceptos como validación, verificación, incertidumbre de medición y trazabilidad, ofrece un lenguaje técnico compartido que tanto los desarrolladores de sistemas de inteligencia artificial forense como los tribunales que deben evaluar esos sistemas pueden utilizar sin ambigüedad, reduciendo el riesgo de que una misma palabra, como validación, se emplee con significados distintos y no siempre rigurosos según quién la utilice, un problema de comunicación que, como se documentó en los capítulos anteriores de este libro a propósito de la terminología de inteligencia artificial, ha contribuido de manera considerable a la confusión que rodea a buena parte del debate público sobre estas tecnologías.

La parte de la serie dedicada al análisis resulta particularmente relevante para este capítulo, porque exige explícitamente que cualquier método forense, y no existe razón alguna para excluir de esa exigencia a un método basado en inteligencia artificial, sea sometido a un proceso documentado de validación antes de su incorporación a la práctica rutinaria, y que ese proceso de validación se revise periódicamente para asegurar que el método sigue siendo válido en las condiciones bajo las cuales efectivamente se aplica. Esta exigencia de revisión periódica adquiere una relevancia particular a la luz de la discusión del capítulo anterior de este libro sobre la velocidad de obsolescencia de los sistemas de inteligencia artificial forense, considerablemente mayor que la de los métodos forenses tradicionales que originalmente motivaron el desarrollo de estos estándares.

La parte de la serie dedicada a la interpretación, por su parte, formula una distinción que resulta directamente aplicable a buena parte de la discusión desarrollada en capítulos anteriores de este libro: la que separa la interpretación investigativa, orientada a generar hipótesis y orientar líneas de investigación durante la etapa de instrucción, de la interpretación evaluativa, orientada a formular una opinión pericial formal que se presentará como prueba ante el tribunal de sentencia. Numerosos sistemas de inteligencia artificial examinados a lo largo de este libro, entre ellos el reconocimiento facial de identificación uno a muchos descrito en el tercer capítulo y los sistemas de análisis predictivo descritos en el capítulo anterior, resultan adecuados, según la evidencia disponible, para cumplir una función investigativa orientadora, pero no satisfacen todavía el estándar de validez exigible a una interpretación evaluativa formal presentada como prueba de cargo. Confundir estos dos niveles de interpretación, tratando como prueba evaluativa lo que en realidad es apenas una orientación investigativa preliminar, constituye uno de los errores de admisibilidad más frecuentes y más consecuentes que este capítulo busca ayudar a prevenir.

Sin embargo, la existencia de un estándar internacional no garantiza, por sí sola, su cumplimiento efectivo por parte de cada laboratorio forense individual, y la acreditación bajo estos estándares sigue siendo, en la mayoría de los países de la región iberoamericana, un proceso voluntario y no una exigencia legal para que un dictamen pericial resulte admisible ante un tribunal. Esta brecha entre la existencia de un estándar técnico riguroso y su exigencia legal efectiva constituye, precisamente, el problema central que las secciones siguientes de este capítulo examinan desde la perspectiva específica de la admisibilidad probatoria.

Conviene señalar, además, que la propia estructura en cinco partes de la serie ISO 21043 ofrece un mapa útil para identificar en qué etapa concreta del proceso forense debería situarse la exigencia de validación de un sistema de inteligencia artificial determinado, una precisión que evita el error, frecuente en el debate público, de tratar la validación algorítmica como una exigencia única y homogénea. Un sistema de inteligencia artificial utilizado en la etapa de reconocimiento y recolección, como un algoritmo de detección automática de evidencia en imágenes de dron, exige un tipo de validación centrado en su capacidad de identificar correctamente elementos relevantes sin generar falsos negativos que hagan perder evidencia crítica. Un sistema utilizado en la etapa de interpretación, como un algoritmo de genotipado probabilístico, exige en cambio una validación centrada en la exactitud estadística de las razones de verosimilitud que produce. Distinguir con precisión en qué etapa opera cada herramienta algorítmica permite formular exigencias de validación específicas y proporcionadas, en lugar de aplicar de manera indiferenciada un mismo checklist genérico a herramientas que cumplen funciones completamente distintas dentro del proceso forense.

8.4. El estándar Daubert y sus factores aplicados a sistemas de inteligencia artificial

El estándar Daubert, establecido por la Corte Suprema de Estados Unidos en 1993 y adoptado, con variaciones, por buena parte de los sistemas de derecho comparado, convirtió al juez en un guardián responsable de evaluar la validez científica de la evidencia experta antes de permitir que llegara al conocimiento del jurado. Los factores tradicionalmente asociados a este estándar, la posibilidad de someter la técnica a prueba empírica, la existencia de revisión por pares y publicación, la tasa de error conocida, la existencia de estándares que controlan la aplicación de la técnica, y su grado de aceptación general dentro de la comunidad científica correspondiente, fueron formulados originalmente pensando en técnicas científicas relativamente estables, no en sistemas de aprendizaje automático que pueden actualizarse, reentrenarse o modificar su comportamiento de manera continua.

Distintos análisis jurídicos recientes coinciden en que la aplicación mecánica de estos cinco factores a un sistema de inteligencia artificial exige adaptaciones específicas, sin las cuales el estándar corre el riesgo de resultar ya sea demasiado permisivo, al admitir sistemas opacos simplemente porque su fabricante afirma haberlos probado internamente, o demasiado restrictivo, al excluir sistemas genuinamente útiles por el solo hecho de no encajar perfectamente en categorías diseñadas para un tipo de evidencia distinto. La tasa de error, por ejemplo, adquiere una complejidad adicional en el contexto de la inteligencia artificial, porque como demostró el capítulo dedicado a los rastros materiales de este libro, esa tasa puede variar de manera considerable según las condiciones específicas de cada caso, lo que exige que los tribunales exijan no una cifra global de exactitud, sino evidencia de que la tasa de error reportada corresponde específicamente a las condiciones del caso concreto que se está juzgando (véase la Figura 8.1).

Factor Daubert	Pregunta tradicional	Exigencia adaptada a un sistema de IA
Verificabilidad empírica	¿Puede someterse la técnica a prueba?	Validación independiente con datos no vistos, no solo pruebas internas del fabricante
Revisión por pares y publicación	¿Fue publicada y revisada por expertos?	Acceso de terceros al modelo y a su documentación pese al secreto comercial
Tasa de error conocida	¿Cuál es su tasa de error?	Tasas desagregadas por población, calidad de la muestra y versión del modelo
Estándares de control	¿Existen estándares que controlen su aplicación?	Protocolos de uso, acreditación (ISO 21043) y control de versiones
Aceptación general	¿La acepta la comunidad científica pertinente?	Respaldo académico independiente, no solo adopción comercial

Figura 8.1. Los cinco factores Daubert y su adaptación a sistemas de inteligencia artificial. La tasa de error, destacada, deja de ser un valor único y debe informarse según población, calidad de la muestra y versión del modelo. Fuente: elaboración propia a partir del análisis desarrollado en esta sección.

Una dimensión adicional del estándar Daubert que merece un tratamiento específico en el contexto de la inteligencia artificial es la responsabilidad legal que recae sobre quienes deciden confiar en un sistema algorítmico sin haber verificado adecuadamente su validez. Un análisis jurídico centrado en lo que denomina discriminación algorítmica negligente sostiene que, así como un profesional puede incurrir en responsabilidad por negligencia al aplicar un método claramente inadecuado a un caso, una institución forense o un operador judicial que confía en un sistema de inteligencia artificial sin haber exigido ni verificado la evidencia de validación que este capítulo ha descrito podría, en determinadas circunstancias, incurrir en una forma análoga de negligencia, particularmente cuando esa confianza ciega produce un resultado discriminatorio o erróneo que una verificación razonable habría podido prevenir (Páez, 2021). Esta perspectiva traslada parte de la responsabilidad de la validación algorítmica desde el desarrollador del sistema, cuya responsabilidad primaria nadie discute, hacia la institución que decide adoptarlo sin las salvaguardas adecuadas, un desplazamiento de responsabilidad que podría incentivar, por la vía de la responsabilidad civil y no solamente de la regulación administrativa, una mayor diligencia institucional en la adopción de estas herramientas.

Una investigación jurídica reciente que examinó decisiones judiciales sobre software forense digital documentó una disparidad considerable entre tribunales a la hora de exigir evidencia rigurosa de validación: mientras algunos tribunales rechazaron software con tasas de error no documentadas y sin respaldo en la comunidad científica, otros admitieron herramientas similares basándose principalmente en la solvencia y el poder de persuasión del testigo experto que las presentó, más que en evidencia empírica independiente de su exactitud. Esta disparidad, documentada dentro de un mismo sistema judicial relativamente homogéneo como el estadounidense, anticipa un desafío todavía mayor para los sistemas judiciales de la región iberoamericana, donde la experiencia acumulada en la aplicación de estándares de admisibilidad científica a la inteligencia artificial es, comparativamente, mucho más incipiente.

Un análisis jurídico centrado específicamente en esta disparidad documenta un caso de 2024 en el que un tribunal federal de apelaciones evaluó la fiabilidad de un software de extracción de datos de localización a partir de un teléfono celular, y contrastó ese caso con otro anterior en

el que un tribunal estatal había rechazado un software de análisis de ubicación por carecer de datos que respaldaran su tasa de error reportada, y por la ausencia de aceptación general de sus métodos dentro de la comunidad científica correspondiente. La comparación entre ambos casos revela un patrón preocupante: cuando la parte proponente de la prueba algorítmica cuenta con un testigo experto convincente, algunos tribunales tienden a aceptar afirmaciones sobre la exactitud del sistema basadas únicamente en la palabra de ese testigo, sin exigir la documentación empírica independiente que el propio estándar Daubert formalmente requiere, un patrón que confirma la necesidad de que los jueces, en su rol de guardianes de la admisibilidad científica, desarrollen una capacitación técnica suficiente para distinguir entre una explicación convincente y una explicación empíricamente respaldada.

8.5. Explicabilidad como requisito de admisibilidad

El segundo capítulo de este libro introdujo el problema de la caja negra y distinguió entre la explicación técnica de un modelo, que identifica qué variables influyeron en una predicción, y la justificación jurídica de una conclusión, que debe satisfacer el estándar argumentativo que exige el derecho probatorio. Esta distinción resulta central para la discusión sobre admisibilidad, porque un análisis jurídico reciente sobre el problema de la caja negra en la adjudicación automatizada sostiene que buena parte de la ansiedad regulatoria en torno a la opacidad algorítmica descansa sobre una comparación implícita, y frecuentemente injusta, entre un estándar de explicabilidad total exigido a la inteligencia artificial y un estándar considerablemente más laxo que, en la práctica, se aplica al razonamiento humano de un examinador forense tradicional, cuyo proceso cognitivo interno resulta, en última instancia, igualmente inaccesible a un escrutinio externo completo (Brożek et al., 2023).

Esta observación no debería leerse como un argumento para relajar las exigencias de explicabilidad de los sistemas de inteligencia artificial forense, sino como una invitación a formular esas exigencias de manera más precisa y menos retórica. Lo que un tribunal necesita de cualquier prueba experta, humana o algorítmica, no es una transparencia absoluta e imposible sobre cada paso del proceso cognitivo o computacional que condujo a una conclusión, sino una justificación suficiente que permita a la parte contraria formular un contrainterrogatorio efectivo y al tribunal evaluar de manera informada la fiabilidad de esa conclusión. Investigaciones específicas sobre informática forense digital proponen, en este sentido, que la explicabilidad se incorpore como requisito de diseño desde el desarrollo inicial de cualquier herramienta destinada a producir prueba forense, en lugar de intentar añadirla como una capa de interpretación posterior sobre un sistema ya opaco (Billah, 2025).

Una consecuencia práctica de este principio es que los tribunales deberían distinguir, al evaluar la explicabilidad de un sistema forense algorítmico, entre dos preguntas que con frecuencia se presentan de manera confusa en un mismo testimonio pericial. La primera pregunta es si el sistema puede explicar por qué llegó a una conclusión determinada en el caso concreto que se juzga, una exigencia de explicabilidad local que las técnicas de interpretabilidad posterior descritas en el segundo capítulo de este libro pueden satisfacer, al menos parcialmente. La segunda pregunta, considerablemente más exigente, es si el sistema en su conjunto puede explicar por qué, en general, produce resultados confiables, una exigencia de explicabilidad global que remite directamente a los estudios de validación fundacional examinados en capítulos anteriores. Un sistema puede satisfacer razonablemente la primera exigencia y fallar en la segunda, o viceversa, y un tribunal que no distinga entre ambas corre el riesgo de conformarse con una explicación local convincente sobre un caso particular, sin haber verificado nunca que el sistema, considerado en su conjunto, merece la confianza que esa explicación local sugiere.

8.6. Cadena de custodia digital en la era algorítmica

La cadena de custodia, el registro documentado de cada persona que tuvo contacto con una pieza de evidencia desde su recolección hasta su presentación ante el tribunal, constituye un requisito procesal tan antiguo como la propia criminalística. La incorporación de sistemas de inteligencia artificial a distintas etapas del proceso forense exige extender esa misma lógica de trazabilidad al ámbito digital y algorítmico, un desafío que la serie ISO 21043 comienza a abordar, pero que plantea preguntas específicas que los protocolos tradicionales de cadena de custodia, diseñados para objetos físicos, no contemplan de manera directa.

Vale la pena ilustrar esta diferencia con un ejemplo concreto que recupera el hilo argumental de capítulos anteriores de este libro. Cuando un examinador dactiloscópico tradicional emite una opinión sobre una coincidencia de huellas, la cadena de custodia debe documentar quién recolectó la huella latente, cómo se transportó y almacenó, y quién realizó la comparación final, pero no necesita documentar el proceso cognitivo interno mediante el cual ese examinador llegó a su conclusión, algo que la propia naturaleza humana del juicio pericial hace imposible de registrar con precisión completa. Cuando ese mismo proceso de comparación se automatiza mediante un sistema de aprendizaje profundo, en cambio, sí resulta técnicamente posible, y por tanto exigible, registrar de manera precisa y reproducible cada parámetro y cada dato de entrada que condujo a la conclusión algorítmica, una posibilidad técnica que no existía para el examinador humano y que, precisamente por existir, se convierte en una obligación de cadena de custodia que ningún laboratorio debería omitir por el solo hecho de que la tecnología tradicional nunca la exigió.

Una cadena de custodia algorítmica completa debería documentar, como mínimo, la versión exacta del software o del modelo utilizado en cada etapa del análisis, dado que un mismo sistema puede comportarse de manera distinta entre una versión y la siguiente, los parámetros específicos de configuración aplicados al caso concreto, cualquier intervención humana en el procesamiento de los datos de entrada, y los resultados intermedios generados en cada etapa del análisis, no únicamente la conclusión final presentada ante el tribunal. Esta última exigencia resulta especialmente relevante para los sistemas de aprendizaje profundo examinados en capítulos anteriores de este libro, donde la diferencia entre el dato bruto capturado, el resultado intermedio de una primera etapa de procesamiento y la conclusión final interpretada puede resultar decisiva para que la defensa pueda ejercer un contrainterrogatorio efectivo sobre cualquiera de esas etapas de manera independiente (véase la Figura 8.2).

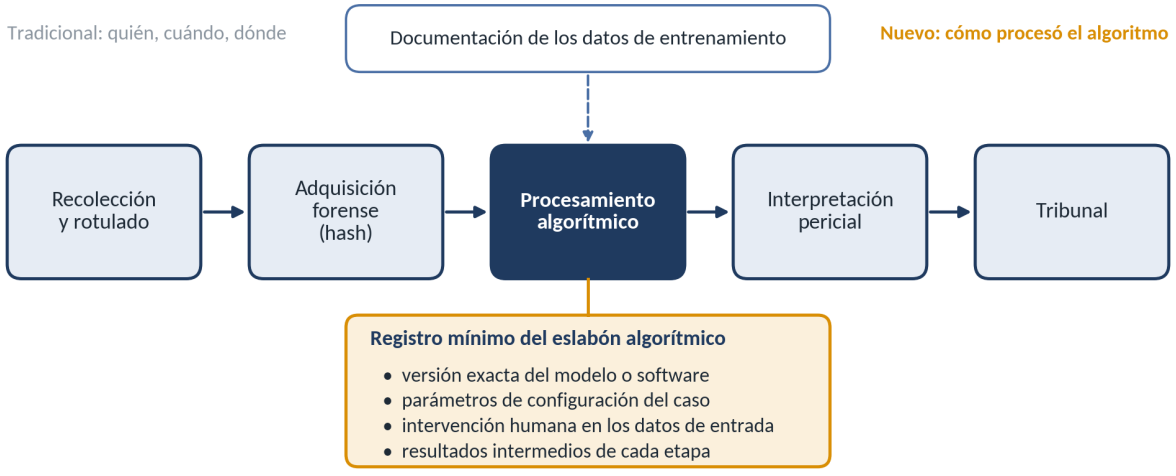


Figura 8.2. Cadena de custodia extendida al eslabón algorítmico. Además de quién, cuándo y dónde, el registro debe permitir reconstruir cómo procesó el algoritmo la evidencia. Fuente: elaboración propia.

Un desafío adicional que la cadena de custodia algorítmica debe resolver, sin equivalente directo en la cadena de custodia física tradicional, es la documentación de los datos de entrenamiento utilizados para construir el modelo, en la medida en que esa información esté disponible y no protegida por secreto comercial. Como se explicó en capítulos anteriores de este libro a propósito de la representatividad poblacional de los conjuntos de entrenamiento, buena parte de las limitaciones de un sistema de inteligencia artificial forense no se manifiestan en el momento del análisis de un caso individual, sino que están inscritas de antemano en la composición del conjunto de datos con el que el modelo aprendió sus patrones. Un protocolo de cadena de custodia algorítmica verdaderamente completo debería, por tanto, exigir que el laboratorio o el proveedor conserve y pueda poner a disposición del tribunal, al menos en sus características generales, una descripción suficiente de esos datos de entrenamiento, de manera que la defensa pueda evaluar si la población representada en ellos se corresponde razonablemente con las características del caso concreto que se juzga.

8.7. La brecha de recursos: acceso a expertos y defensa efectiva

Un problema estructural que atraviesa la admisibilidad de la prueba algorítmica, y que la literatura jurídica ha comenzado a documentar con particular preocupación, es la asimetría de recursos entre la fiscalía, que habitualmente cuenta con acceso institucional a las herramientas de inteligencia artificial forense y a la capacitación necesaria para presentarlas ante un tribunal, y la defensa, particularmente la defensa pública que representa a personas sin recursos económicos propios, que con frecuencia carece del acceso a peritos independientes capaces de cuestionar de manera efectiva la validez de esa misma prueba algorítmica. Un análisis reciente centrado específicamente en esta asimetría documenta casos en los que la defensa se vio obligada a aceptar sin cuestionamiento efectivo los resultados de software forense digital de origen comercial, precisamente por no contar con los recursos necesarios para contratar un perito independiente capaz de auditar ese software (Melkote y Nambiar, 2025).

Este mismo análisis documenta un caso ilustrativo en el que un tribunal de apelaciones confirmó la admisión de evidencia de localización extraída mediante software forense de un teléfono celular, presentada a través del testimonio de un agente de las fuerzas de seguridad que actuó como testigo lego y no como perito propiamente acreditado, sin que la defensa hubiera podido cuestionar de manera efectiva la metodología subyacente del software utilizado (Melkote y Nambiar, 2025). Este tipo de escenario, en el cual una herramienta algorítmica sofisticada llega al conocimiento del tribunal a través de un testimonio que no satisface plenamente los estándares exigibles a la prueba pericial científica, ilustra un riesgo que ningún marco de admisibilidad, por riguroso que sea en su formulación teórica, puede prevenir si las partes del proceso no cuentan con recursos equivalentes para cuestionarlo. La garantía de igualdad de armas entre acusación y defensa, principio fundamental de cualquier proceso penal justo, adquiere en la era de la prueba algorítmica una dimensión técnica y económica que los sistemas de justicia de la región iberoamericana, muchos de ellos con defensorías públicas estructuralmente subfinanciadas, deberían considerar con particular atención.

El mismo análisis propone, como respuesta institucional concreta a esta asimetría, reconocer un derecho explícito de las personas acusadas sin recursos propios a contar con asistencia pericial independiente costeada por el Estado cuando la acusación se sustente de manera sustancial en prueba algorítmica, un derecho que se derivaría de garantías constitucionales más generales de defensa efectiva ya reconocidas en la mayoría de los sistemas jurídicos de la región, pero que rara vez se ha articulado de manera específica para el contexto de la prueba basada en inteligencia artificial (Melkote y Nambiar, 2025). Una alternativa institucional complementaria, particularmente

viable para países con recursos limitados que no pueden financiar peritos independientes caso por caso, es la creación de laboratorios de auditoría algorítmica de carácter público y centralizado, capaces de prestar servicio a las defensorías públicas de todo un país o de toda una región, en lugar de exigir que cada defensor individual busque, caso por caso, un perito privado dispuesto y capaz de auditar un sistema algorítmico específico. Este modelo centralizado, todavía inexistente en la mayoría de los países de la región iberoamericana, podría representar una solución más eficiente en términos de costo que la alternativa de multiplicar la capacidad pericial independiente en cada jurisdicción de manera aislada.

8.8. Sistemas acusatorios latinoamericanos: entre la aceptación general y el escrutinio Daubert

Como se anticipó en el primer capítulo de este libro, la reforma procesal penal que trasladó a la mayoría de los países latinoamericanos desde modelos inquisitivos hacia sistemas de corte acusatorio elevó considerablemente la relevancia de la prueba pericial presentada en juicio oral, sin que ese cambio procesal haya venido acompañado, en la mayoría de los casos, de un desarrollo equivalente de filtros de admisibilidad específicamente diseñados para evaluar la validez científica de esa prueba. Un análisis comparado reciente sobre la exclusión de prueba pericial de baja calidad epistémica documenta que buena parte de los sistemas procesales de tradición continental, incluida la mayoría de los países latinoamericanos, carecen todavía de un momento procesal claramente definido y de criterios epistémicos explícitos para excluir, antes del juicio oral, una prueba pericial que no satisfaga estándares mínimos de validez científica (Sánchez, 2021).

Frente a esta laguna, distintas instituciones forenses de la región han comenzado a incorporar, por iniciativa propia y sin necesariamente contar con respaldo legislativo expreso, criterios inspirados directamente en el estándar Daubert estadounidense como guía de buenas prácticas internas. El Instituto Nacional de Ciencias Forenses de Guatemala, por ejemplo, ha documentado la adopción de un sistema de gestión de calidad que incorpora explícitamente los criterios de la sentencia Daubert como referencia para sus propios procedimientos periciales, en ausencia de una exigencia legal formal equivalente en el derecho procesal guatemalteco (García Morales, 2021). Iniciativas similares, documentadas en México y en otros países de la región, buscan ofrecer a los jueces, que como se señaló anteriormente en este capítulo no cuentan típicamente con formación científica especializada, lineamientos concretos para valorar la confiabilidad de la prueba pericial en disciplinas como la genética forense, la dactiloscopia y el análisis de voz, entre otras.

Estas iniciativas de adopción institucional voluntaria comparten, sin embargo, una limitación estructural que conviene señalar con claridad: al no contar con respaldo legislativo formal, su cumplimiento efectivo depende enteramente de la voluntad y de los recursos de cada institución particular, sin que exista un mecanismo de exigibilidad judicial que obligue a un laboratorio a aplicarlas, ni una consecuencia procesal clara para el incumplimiento. Un dictamen pericial elaborado sin seguir estos lineamientos voluntarios sigue siendo, en la mayoría de los sistemas procesales de la región, formalmente admisible ante un tribunal, precisamente porque la ausencia de un filtro legislativo obligatorio de admisibilidad científica, ya señalada en el párrafo anterior, significa que ningún juez está legalmente obligado a excluir esa prueba únicamente por no haber seguido las buenas prácticas voluntarias que la propia comunidad forense se ha autoimpuesto.

Estas iniciativas, valiosas como respuesta institucional parcial, no sustituyen la necesidad de una reforma legislativa explícita que incorpore, en los códigos procesales penales de la región, un filtro de admisibilidad formal y obligatorio para la prueba científica y algorítmica, equivalente en su función, aunque no necesariamente idéntico en su formulación exacta, al estándar Daubert

estadounidense. Sin ese respaldo legislativo, la adopción de buenas prácticas de validación queda librada a la voluntad institucional de cada laboratorio forense particular, una situación de fragilidad normativa que, precisamente en el momento en que la región comienza a incorporar herramientas de inteligencia artificial a su práctica forense, según documentó cada uno de los capítulos anteriores de este libro, resulta especialmente riesgosa.

Esta reforma legislativa pendiente adquiere una urgencia particular si se considera la velocidad de adopción de herramientas de inteligencia artificial documentada a lo largo de este libro, considerablemente mayor que la velocidad histórica con la que la región logró implementar la propia reforma procesal penal acusatoria, un proceso que en varios países tomó más de una década desde su aprobación legislativa hasta su plena implementación operativa. Si la región reprodujera ese mismo ritmo pausado de reforma institucional para incorporar un filtro de admisibilidad específico para la prueba algorítmica, correría el riesgo de que buena parte de esa tecnología ya se encuentre profundamente integrada en la práctica forense cotidiana, con años de dictámenes periciales ya presentados ante los tribunales, para cuando finalmente exista un marco legal capaz de someterla al escrutinio que este capítulo ha defendido. Anticipar esa reforma, en lugar de esperar a que una crisis de validez análoga a la descrita en el primer capítulo de este libro obligue a implementarla de manera reactiva, constituye quizás la recomendación de política pública más urgente que se desprende de todo este capítulo.

8.9. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la aplicación mecánica de estándares de admisibilidad diseñados para tecnologías estables a sistemas de inteligencia artificial de actualización continua, la exigencia de una explicabilidad algorítmica desproporcionadamente más estricta que la que se aplica al razonamiento humano de un examinador tradicional, la ausencia de protocolos de cadena de custodia adaptados a las particularidades del procesamiento algorítmico, la asimetría de recursos técnicos entre fiscalía y defensa para cuestionar la prueba algorítmica, y la ausencia, en buena parte de la región iberoamericana, de un filtro de admisibilidad legislativo formal para la prueba científica y algorítmica.

Un riesgo adicional, de carácter más procesal que técnico, merece destacarse antes de cerrar este apartado. La ausencia de un momento procesal claramente delimitado para discutir la admisibilidad científica de una prueba, un problema que la literatura comparada sobre sistemas de tradición continental identifica como una laguna estructural compartida por buena parte de los países de la región, significa que la validez de un sistema de inteligencia artificial forense puede terminar discutiéndose, si es que se discute en absoluto, recién durante el juicio oral, ante el propio tribunal de sentencia, en lugar de resolverse en una audiencia previa específicamente dedicada a esa cuestión, como ocurre bajo el modelo de audiencia Daubert del sistema estadounidense. Esta diferencia procesal no es meramente formal: una discusión de admisibilidad científica resuelta antes del juicio oral permite excluir prueba inválida sin que el tribunal de sentencia llegue siquiera a conocerla, mientras que una discusión postergada hasta el juicio oral expone al juzgador a una prueba que, aunque eventualmente se declare inadmisible o se le reste valor, ya habrá ejercido una influencia psicológica difícil de revertir por completo, un riesgo que la psicología del razonamiento judicial ha documentado de manera consistente en otros contextos.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir que cualquier sistema

de inteligencia artificial forense documente su tasa de error específicamente para las condiciones del caso concreto y no únicamente una cifra global genérica, incorporar la explicabilidad como requisito de diseño desde el desarrollo de cualquier herramienta forense algorítmica, extender los protocolos de cadena de custodia para documentar versión, configuración e intervención humana en cada etapa del procesamiento algorítmico, garantizar el acceso efectivo de la defensa, incluida la defensa pública, a peritos independientes capaces de auditar la prueba algorítmica presentada por la fiscalía, e impulsar reformas legislativas que incorporen un filtro de admisibilidad formal para la prueba científica y algorítmica en los sistemas procesales de la región.

A estas prácticas conviene sumar una recomendación de cierre que sintetiza el argumento central de este capítulo y, en buena medida, de todo este libro. Ningún avance técnico en la validación estadística, en la explicabilidad algorítmica o en la documentación de la cadena de custodia puede sustituir la exigencia jurídica de fondo que este capítulo ha desarrollado: que toda prueba presentada ante un tribunal, sin importar si fue generada por un examinador humano o por un sistema de inteligencia artificial, debe superar un escrutinio genuino sobre su validez científica antes de poder influir en la libertad de una persona. La inteligencia artificial no exime a la criminalística de esa exigencia histórica; la actualiza, la acelera y, en el mejor de los casos, ofrece herramientas metodológicas que permiten satisfacerla con un rigor que la ciencia forense tradicional tardó más de un siglo en alcanzar, según documentó el primer capítulo de este libro.



CAPÍTULO 9

Ética, sesgo y derechos humanos: sesgo algorítmico, privacidad y marcos regulatorios

9.1. Caso de apertura: Machine Bias y el nacimiento del debate sobre sesgo algorítmico en la justicia penal

En mayo de 2016, el medio periodístico de investigación ProPublica publicó un análisis que se convertiría en el punto de partida obligado de prácticamente toda la discusión posterior sobre sesgo algorítmico en el sistema de justicia penal. El reportaje, titulado Machine Bias, examinó el funcionamiento de COMPAS, un software comercial y propietario utilizado en numerosas jurisdicciones estadounidenses para estimar la probabilidad de que una persona acusada reincidiera, una estimación que jueces y funcionarios de libertad condicional utilizaban para tomar decisiones sobre fianza, sentencia y supervisión (Angwin et al., 2016).

Los periodistas reunieron datos de más de diez mil personas acusadas en el condado de Broward, Florida, y compararon las puntuaciones de riesgo que COMPAS había asignado a cada una de ellas con lo que efectivamente ocurrió durante los dos años siguientes. El hallazgo central resultó estadísticamente contundente: entre las personas que finalmente no reincidieron, los acusados de raza negra tuvieron casi el doble de probabilidad de haber sido clasificados erróneamente como de alto riesgo, en comparación con los acusados de raza blanca, cuarenta y cinco por ciento frente a veintitrés por ciento. De manera inversa, entre quienes sí reincidieron, los acusados blancos fueron clasificados erróneamente como de bajo riesgo con casi el doble de frecuencia que los acusados negros (Angwin et al., 2016).

Conviene señalar, para situar correctamente el alcance de este hallazgo, que el propio análisis de ProPublica documentó también un dato adicional que rara vez se menciona con la misma frecuencia que las disparidades raciales: la exactitud predictiva general del sistema COMPAS resultó, en términos absolutos, considerablemente modesta, con una capacidad de predecir correctamente la reincidencia general de apenas el sesenta y un por ciento de los casos, y de la reincidencia violenta específicamente de solo alrededor del veinte por ciento. Esta combinación de exactitud predictiva limitada y disparidad demográfica sistemática resulta particularmente inquietante cuando se considera que decisiones judiciales con consecuencias directas sobre la libertad de las personas, como la fijación de una fianza o la duración de una supervisión de libertad condicional, se apoyaban, y en muchas jurisdicciones estadounidenses todavía se apoyan, en un instrumento cuya capacidad predictiva apenas supera, en el mejor de los casos, a un lanzamiento de moneda ligeramente sesgado.

La empresa desarrolladora del software, Northpointe, respondió que su sistema era igualmente preciso para ambos grupos raciales cuando se medía mediante una métrica distinta, la calibración, es decir, la correspondencia entre la puntuación de riesgo asignada y la probabilidad real de reincidencia dentro de cada grupo. Investigaciones académicas posteriores demostraron que ambas partes tenían razón, en un sentido técnico preciso: es matemáticamente imposible que un sistema de predicción satisfaga simultáneamente la métrica de equidad que defendía ProPublica y la métrica de calibración que defendía Northpointe, cuando la tasa base de reincidencia difiere entre los grupos comparados, un resultado conocido en la literatura técnica como el teorema de imposibilidad de la equidad algorítmica. Este descubrimiento, que este capítulo retomará con mayor detalle más adelante, transformó lo que parecía una disputa resoluble mediante mejor ingeniería en un dilema genuinamente normativo, que ninguna solución puramente técnica puede resolver por sí sola.

Conviene precisar, para comprender cabalmente el alcance de este dilema, qué significa exactamente que dos métricas de equidad resulten matemáticamente incompatibles entre sí. La métrica que privilegiaba ProPublica, conocida técnicamente como paridad de errores, exige que la proporción de falsos positivos y falsos negativos sea igual entre los grupos comparados, de manera

que ningún grupo cargue de manera desproporcionada con el costo de una clasificación errónea. La métrica que privilegiaba Northpointe, la calibración, exige en cambio que una puntuación de riesgo determinada signifique lo mismo, en términos de probabilidad real de reincidencia, sin importar el grupo racial del individuo evaluado. Cuando la tasa de reincidencia real difiere entre grupos, como efectivamente ocurría en los datos de Broward County por razones que remiten a las mismas dinámicas estructurales de vigilancia desigual discutidas en el capítulo anterior de este libro, resulta matemáticamente imposible satisfacer ambas exigencias de manera simultánea, un resultado formalizado independientemente por varios equipos de investigación en ciencia de datos poco después de la publicación original de ProPublica, y que hoy constituye uno de los hallazgos más citados de toda la literatura sobre equidad algorítmica.

9.2. De la validez técnica a la justicia algorítmica

Los ocho capítulos anteriores de este libro se han concentrado, en su gran mayoría, en una pregunta de validez técnica: si un sistema de inteligencia artificial forense produce resultados exactos, confiables y explicables. Este capítulo final del bloque dedicado a los desafíos de la inteligencia artificial forense se ocupa de una pregunta distinta, aunque estrechamente relacionada: incluso cuando un sistema es técnicamente válido en el sentido desarrollado en capítulos anteriores, ¿puede su uso resultar, pese a ello, éticamente problemático o violatorio de derechos fundamentales? El caso de apertura de este capítulo demuestra que la respuesta es afirmativa, y que la validez técnica y la justicia algorítmica constituyen dos exigencias relacionadas pero no idénticas, que la criminalística del futuro debe aprender a evaluar de manera conjunta y no sucesiva.

Esta distinción entre validez técnica y justicia algorítmica merece una ilustración adicional que ayude a fijarla con claridad. Un sistema de evaluación de riesgo de reincidencia podría, en principio, alcanzar niveles de exactitud predictiva considerablemente más altos que los documentados para COMPAS, y superar sin dificultad cualquier estudio de validación fundacional del tipo exigido en capítulos anteriores de este libro, y aun así perpetuar o incluso amplificar disparidades raciales profundamente arraigadas en la historia de la vigilancia policial de una sociedad determinada, precisamente porque esa exactitud técnica se calcula sobre datos que ya incorporan, de manera invisible, las consecuencias acumuladas de décadas de aplicación desigual de la ley. La validez técnica, en otras palabras, es una condición necesaria pero no suficiente para la justicia algorítmica, un principio que ningún capítulo anterior de este libro debería llevar a subestimar, y que este capítulo final se propone corregir de manera explícita.

Este capítulo examina tres dimensiones de ese problema más amplio de justicia algorítmica: el sesgo algorítmico propiamente dicho, ya introducido en capítulos anteriores a propósito de disciplinas específicas y que aquí se sistematiza de manera transversal; la privacidad y la vigilancia biométrica, con especial atención al marco regulatorio más desarrollado hasta la fecha, el Reglamento de Inteligencia Artificial de la Unión Europea; y el estado incipiente, pero creciente, de los marcos regulatorios específicos para la inteligencia artificial en la región iberoamericana.

Conviene anticipar, antes de desarrollar cada una de estas tres dimensiones, la razón por la cual este libro las reserva para su capítulo final dentro del bloque de desafíos, en lugar de dispersarlas a lo largo de cada capítulo temático anterior. Aunque el sesgo demográfico, la privacidad y la ausencia de marcos regulatorios aparecieron ya, de manera puntual, en capítulos dedicados a disciplinas específicas como la identificación biométrica o el análisis predictivo, tratarlos de manera unificada en un capítulo propio permite mostrar que no se trata de problemas aislados de una u otra tecnología particular, sino de un desafío estructural que atraviesa a la criminalística asistida por inteligencia artificial en su conjunto, y que exige, por tanto, una respuesta institucional igualmente transversal,

y no una serie de soluciones fragmentadas y específicas para cada disciplina examinada por separado.

9.3. Sesgo algorítmico: tipos, fuentes y el dilema de las métricas de equidad en competencia

El sesgo algorítmico puede originarse en al menos tres momentos distintos del ciclo de vida de un sistema de inteligencia artificial, una distinción que resulta indispensable para diagnosticar correctamente su origen y, en consecuencia, para diseñar una respuesta adecuada. El sesgo de datos surge cuando el conjunto de entrenamiento no representa de manera adecuada a la población sobre la cual el sistema se aplicará posteriormente, el problema central detrás del desempeño demográfico diferencial en reconocimiento facial documentado en el segundo y el tercer capítulo de este libro (Grother et al., 2019). El sesgo de medición surge cuando la propia variable que el sistema intenta predecir, como la reincidencia en el caso COMPAS, se mide de manera imperfecta y sistemáticamente distinta entre grupos, por ejemplo porque la vigilancia policial más intensa sobre determinados barrios genera más arrestos registrados sin que ello refleje necesariamente más actividad delictiva real, un problema ya discutido en el capítulo anterior de este libro a propósito del análisis predictivo. El sesgo algorítmico propiamente dicho, finalmente, surge de decisiones de diseño del propio modelo, como la elección de qué variables incluir o qué función de pérdida optimizar, decisiones que pueden introducir o amplificar disparidades incluso cuando los datos de entrada fueran, en principio, representativos.

El dilema técnico revelado por el caso COMPAS, la imposibilidad matemática de satisfacer simultáneamente distintas métricas de equidad cuando las tasas base difieren entre grupos, no constituye un defecto corregible mediante mejor programación, sino una limitación estructural que obliga a quien diseña o adopta un sistema de este tipo a tomar una decisión explícitamente normativa sobre qué noción de equidad priorizar en cada contexto de aplicación específico. Esta decisión, lejos de ser un detalle técnico que pueda delegarse enteramente a un equipo de ingeniería, constituye una decisión de política pública con consecuencias directas sobre la libertad de las personas, y como tal debería someterse a un proceso de deliberación pública, jurídica y no meramente técnica, antes de que cualquier sistema de este tipo se incorpore a la práctica de un sistema de justicia penal (véase la Figura 9.1).

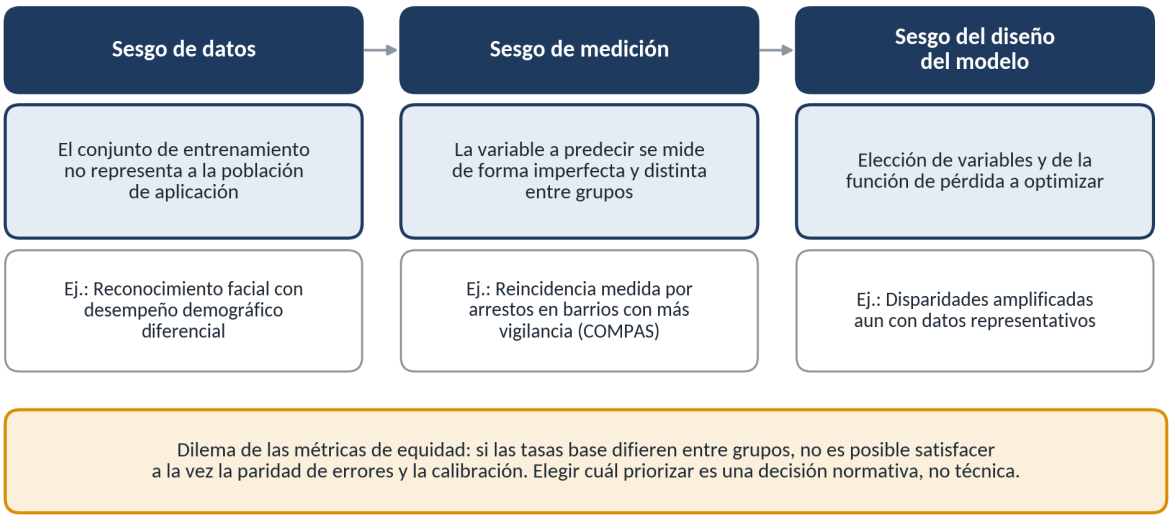


Figura 9.1. Tres fuentes de sesgo algorítmico y el dilema de las métricas de equidad. Diagnosticar en qué momento se origina el sesgo determina qué respuesta resulta adecuada. Fuente: elaboración propia a partir de Grother et al. (2019) y del análisis del caso COMPAS.

Además de la paridad de errores y la calibración ya descritas, la literatura técnica sobre equidad algorítmica ha propuesto docenas de definiciones matemáticas adicionales de lo que constituye un sistema justo, entre ellas la paridad demográfica, que exige que la proporción de resultados positivos sea igual entre grupos sin importar la tasa base real, y la equidad contrafactual, que exige que la predicción para un individuo determinado no cambiaría si ese individuo hubiera pertenecido a un grupo demográfico distinto, manteniendo todo lo demás constante. Cada una de estas definiciones captura una intuición moral genuina y defendible sobre lo que significa tratar a las personas de manera justa, y cada una de ellas, como demostraron los mismos trabajos que formalizaron el teorema de imposibilidad mencionado en el párrafo anterior, resulta generalmente incompatible con las demás en presencia de tasas base distintas entre grupos. Esta proliferación de definiciones técnicas, lejos de ofrecer una solución, ilustra la magnitud del desafío normativo que este capítulo describe: no existe, ni puede existir, una definición matemática única y neutral de equidad que resuelva de una vez y para siempre la pregunta de qué significa que un sistema forense trate a las personas de manera justa.

Conviene precisar que el sesgo algorítmico, en el contexto específico de la criminalística examinado a lo largo de este libro, no se limita a los sistemas de evaluación de riesgo como COMPAS, sino que atraviesa prácticamente todas las disciplinas descritas en capítulos anteriores: el desempeño demográfico diferencial del reconocimiento facial descrito en el tercer capítulo, la escasa representación de poblaciones latinoamericanas en los modelos de antropología y odontología forense descrita en ese mismo capítulo, y los sesgos de retroalimentación del análisis predictivo descritos en el capítulo anterior constituyen, todos ellos, manifestaciones específicas del mismo problema estructural que este capítulo sistematiza de manera transversal.

Resulta importante distinguir, además, entre el sesgo algorítmico y la discriminación algorítmica en sentido jurídico, dos conceptos relacionados pero no idénticos que el debate público frecuentemente confunde. Un sistema puede exhibir una disparidad estadística medible entre grupos, lo que técnicamente constituye sesgo en el sentido descrito en los párrafos anteriores, sin que esa disparidad configure necesariamente una discriminación ilegal en el sentido que los marcos jurídicos de derechos humanos reconocen, si esa disparidad responde a una diferencia real y objetivamente justificada en la variable que el sistema intenta medir. La dificultad práctica, que el propio caso COMPAS ilustra con particular claridad, es que determinar si una disparidad estadística observada refleja una diferencia real subyacente o, por el contrario, un sesgo de medición heredado de prácticas históricas de vigilancia desigual, exige un análisis causal considerablemente más sofisticado que la simple comparación de tasas de error entre grupos, y frecuentemente ese análisis causal no puede resolverse únicamente con los datos disponibles dentro del propio sistema de justicia penal, sino que requiere información social y demográfica más amplia que rara vez se recopila de manera sistemática.

Esta dificultad de atribución causal tiene una consecuencia práctica que las instituciones forenses de la región deberían internalizar como principio de precaución: ante la imposibilidad frecuente de determinar con certeza si una disparidad estadística observada en un sistema algorítmico refleja sesgo de medición o diferencia real, la carga de la prueba debería recaer sobre quien pretende adoptar o mantener en uso el sistema en cuestión, y no sobre quienes, potencialmente afectados por esa disparidad, deben demostrar de manera afirmativa que sufrieron una discriminación. Este principio de precaución, coherente con el enfoque general de gestión de riesgo que adopta el reglamento europeo examinado en la siguiente sección de este capítulo, desplaza el punto de partida del debate: en lugar de asumir que un sistema es justo hasta que se demuestre lo contrario, propone asumir que cualquier disparidad demográfica no explicada constituye una señal de alarma que exige investigación activa antes de continuar utilizando el sistema, un estándar considerablemente

más exigente que el que rige hoy la práctica de la mayoría de las instituciones forenses examinadas a lo largo de este libro.

9.4. Privacidad y vigilancia biométrica: el marco europeo como referencia regulatoria

El Reglamento de Inteligencia Artificial de la Unión Europea, que entró en vigor el primero de agosto de 2024 con sus disposiciones sobre prácticas prohibidas exigibles desde febrero de 2025, constituye hasta la fecha el marco regulatorio más desarrollado y más directamente aplicable a la criminalística de todos los examinados en este libro, precisamente porque dedica una atención específica y detallada a los usos de inteligencia artificial en el contexto de la investigación y la persecución penal (Parlamento Europeo y Consejo de la Unión Europea, 2024).

El reglamento clasifica los sistemas de inteligencia artificial según cuatro niveles de riesgo, y prohíbe de manera categórica un conjunto reducido de prácticas consideradas incompatibles con los derechos fundamentales, entre ellas la evaluación individual del riesgo de que una persona cometa un delito basada exclusivamente en la elaboración de perfiles o en rasgos de personalidad, sin sustento en hechos objetivos verificables, una prohibición que alcanza de manera directa a las modalidades de análisis predictivo basado en personas que el capítulo anterior de este libro identificó como las más problemáticas desde la perspectiva de los derechos fundamentales. El reglamento prohíbe también la creación de bases de datos de reconocimiento facial mediante extracción masiva e indiscriminada de imágenes de internet o de circuitos cerrados de televisión, una prohibición formulada en respuesta directa a las prácticas de empresas como Clearview AI, cuya tecnología de reconocimiento facial construida sobre miles de millones de imágenes extraídas sin consentimiento de plataformas públicas generó sanciones regulatorias considerables en distintas jurisdicciones europeas.

Resulta ilustrativo detenerse brevemente en el caso Clearview AI mencionado en el párrafo anterior, porque conecta de manera directa con la discusión sobre genealogía genética investigativa desarrollada en el cuarto capítulo de este libro. Al igual que ocurrió con las plataformas de genealogía genética de uso recreativo examinadas en ese capítulo, Clearview AI construyó su base de datos de reconocimiento facial recopilando información que millones de personas habían compartido públicamente en redes sociales con un propósito completamente distinto, ajeno a cualquier finalidad de investigación policial, y sin haber solicitado jamás su consentimiento para ese uso secundario. Distintas autoridades europeas de protección de datos, así como reguladores de otras jurisdicciones, impusieron a la empresa sanciones económicas considerables y órdenes de eliminación de datos por esta práctica, un patrón regulatorio que confirma una tesis que atraviesa buena parte de este libro: la reutilización no consentida de datos personales recopilados originalmente con un propósito ajeno a la investigación criminal constituye, independientemente de la disciplina forense específica de que se trate, una de las fuentes más recurrentes de controversia ética y legal en la criminalística asistida por inteligencia artificial.

La identificación biométrica remota en tiempo real en espacios públicos con fines de aplicación de la ley, la aplicación más directamente relevante para la investigación criminal examinada en el tercer capítulo de este libro, queda prohibida como regla general, aunque el reglamento contempla excepciones estrechamente delimitadas para la búsqueda de víctimas específicas de delito, la prevención de amenazas concretas y sustanciales a la seguridad pública, y la identificación de sospechosos de delitos graves específicamente enumerados. Crucialmente, el uso de cualquiera de estas excepciones exige autorización judicial o administrativa independiente previa, salvo en situaciones de urgencia debidamente justificadas, y una evaluación de impacto sobre derechos

fundamentales documentada y registrada ante las autoridades correspondientes (Parlamento Europeo y Consejo de la Unión Europea, 2024).

Conviene precisar, para evitar una lectura excesivamente optimista de este régimen, que la propia implementación práctica de estas prohibiciones ha generado ya, según reportan distintos análisis de seguimiento publicados durante los primeros meses de vigencia efectiva del reglamento, una brecha considerable entre la letra de la norma y su cumplimiento real por parte de las fuerzas de seguridad de distintos países europeos. Las excepciones que permiten el uso de identificación biométrica remota para la persecución de delitos graves se han formulado, en la práctica administrativa de algunas agencias, de manera lo suficientemente amplia como para cubrir buena parte de los escenarios de investigación criminal habituales, lo que ha llevado a organizaciones de derechos digitales a advertir que la prohibición general, pese a representar sobre el papel la restricción legislativa más significativa jamás impuesta a una tecnología de vigilancia específica, corre el riesgo de convertirse, en la práctica cotidiana, en una regla con más excepciones que aplicaciones efectivas. Esta brecha entre la norma escrita y su cumplimiento real ofrece una lección adicional, más allá del contenido sustantivo del propio reglamento: ninguna regulación, por rigurosa que resulte en su formulación, garantiza por sí sola un cambio efectivo en la práctica institucional si no se acompaña de mecanismos robustos de supervisión, auditoría y sanción efectiva por incumplimiento.

Este diseño regulatorio, que condiciona el uso excepcional de una tecnología de por sí prohibida a una autorización judicial previa y a una evaluación de impacto documentada, ofrece un modelo directamente trasladable a la discusión sobre genealogía genética investigativa que el cuarto capítulo de este libro desarrolló en profundidad, y confirma que la lógica de autorización judicial previa como salvaguarda frente a tecnologías de alto riesgo para los derechos fundamentales no es exclusiva de un dominio tecnológico particular, sino un principio de gobernanza que se repite de manera consistente en la experiencia comparada más avanzada, tanto para la biometría facial como para la genética forense de nueva generación.

Más allá de las prohibiciones y de la identificación biométrica remota, el reglamento europeo clasifica como de alto riesgo, sujetos por tanto a obligaciones considerablemente más estrictas que las aplicables a un sistema de riesgo limitado o mínimo, a los sistemas de inteligencia artificial destinados a evaluar la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de un delito, a los sistemas destinados a evaluar el riesgo de que una persona sea víctima de un delito, y a los sistemas de categorización biométrica utilizados para inferir o deducir características sensibles como la etnia o las opiniones políticas. Esta clasificación de alto riesgo, aplicada a buena parte de las tecnologías forenses examinadas a lo largo de este libro, entre ellas los sistemas de genotipado probabilístico del cuarto capítulo y los sistemas de interpretación algorítmica de puntajes balísticos del sexto capítulo, exige a sus desarrolladores y a las instituciones que los adoptan cumplir un conjunto de obligaciones de gestión de riesgo, gobernanza de datos, documentación técnica, supervisión humana y ciberseguridad que, en conjunto, constituyen el intento regulatorio más ambicioso hasta la fecha de trasladar a la ley las mismas exigencias de validación que este libro ha defendido desde su primer capítulo por razones puramente epistemológicas y no solo jurídicas.

Un aspecto del reglamento europeo que merece destacarse por su relevancia directa para las instituciones forenses es la obligación de alfabetización en inteligencia artificial que impone tanto a proveedores como a usuarios de sistemas de alto riesgo, una obligación que exige garantizar un nivel suficiente de comprensión técnica entre el personal que opera o supervisa estos sistemas, proporcional a su contexto de uso. Trasladada al ámbito forense examinado en este libro, esta obligación confirma y refuerza una recomendación que ha aparecido de manera recurrente en capítulos anteriores, particularmente en el segundo capítulo dedicado a los fundamentos de la

inteligencia artificial y en el capítulo anterior sobre admisibilidad probatoria: ningún perito, fiscal, defensor o juez debería operar o evaluar un sistema de inteligencia artificial forense sin haber alcanzado antes un nivel mínimo de alfabetización técnica sobre sus fundamentos, sus limitaciones y sus condiciones de validación, una exigencia que el capítulo final de este libro, dedicado al perito del futuro, retomará con el detalle formativo que merece.

9.5. Marcos regulatorios en América Latina: entre la aspiración y la implementación

A diferencia del marco europeo descrito en la sección anterior, la región iberoamericana se encuentra todavía en una etapa considerablemente más temprana de desarrollo regulatorio específico para la inteligencia artificial. Un relevamiento regional publicado a comienzos de 2024 documenta que, si bien numerosos países de la región han iniciado procesos legislativos o han publicado estrategias nacionales de inteligencia artificial, la mayoría de esas iniciativas todavía carecen de la especificidad normativa y de los mecanismos de exigibilidad que caracterizan al marco europeo (Access Now, 2024).

Este relevamiento regional identifica, además, un patrón que conviene destacar por su relevancia directa para el argumento de este libro: buena parte de las iniciativas legislativas sobre inteligencia artificial actualmente en discusión en distintos países de la región no contienen, hasta la fecha, ninguna disposición específica dedicada al uso de estas tecnologías en el contexto de la investigación criminal o de la administración de justicia penal, un vacío notable si se considera que precisamente ese es el ámbito de aplicación que el reglamento europeo trata con mayor detalle y mayor exigencia normativa, según se describió en la sección anterior de este capítulo. Esta omisión sugiere que el debate legislativo regional sobre inteligencia artificial se ha concentrado, hasta ahora, predominantemente en aplicaciones económicas y comerciales, dejando en un segundo plano precisamente el ámbito de aplicación, la criminalística forense, que constituye el objeto central de este libro y que, por las razones desarrolladas a lo largo de sus nueve capítulos, exige una atención regulatoria al menos tan urgente como la que reciben las aplicaciones comerciales de esta misma tecnología.

Perú se convirtió en 2023 en el primer país de la región en sancionar una ley específica sobre inteligencia artificial, aunque se trata de una norma declarativa de apenas cinco artículos que proclama principios generales sin establecer todavía mecanismos operativos concretos de supervisión o de sanción. Brasil, por su parte, avanza en la discusión legislativa de un proyecto de ley que sigue de manera relativamente cercana el enfoque de clasificación por niveles de riesgo del reglamento europeo, aunque su aprobación definitiva seguía pendiente al momento de redactarse este libro. Otros países de la región cuentan con iniciativas más acotadas, centradas en sectores específicos como el financiero, o con mesas de trabajo interministeriales orientadas todavía a la fase de diagnóstico y diseño de política pública, más que a la sanción de normas jurídicamente exigibles.

Esta heterogeneidad regulatoria entre los países de la región, documentada de manera sistemática por el relevamiento citado en el párrafo anterior, contrasta con la relativa homogeneidad que caracterizó, dos décadas atrás, a la reforma procesal penal acusatoria descrita en el primer capítulo de este libro, un proceso que, aunque desigual en sus tiempos, siguió un patrón regional razonablemente coordinado, apoyado por organismos de cooperación judicial y por el intercambio activo de experiencias entre los distintos países. La ausencia, hasta la fecha, de un esfuerzo de coordinación regional equivalente en materia de regulación de inteligencia artificial forense representa una oportunidad perdida, particularmente porque numerosos desafíos examinados a lo

largo de este libro, desde la representatividad poblacional de los modelos de antropología forense hasta la gobernanza de la genealogía genética investigativa, comparten características estructurales comunes entre los distintos países de la región, y podrían beneficiarse de estándares armonizados que faciliten tanto la cooperación judicial transfronteriza como el aprovechamiento compartido de recursos técnicos escasos.

Esta asimetría regulatoria entre Europa y la región iberoamericana plantea un riesgo que conviene formular con claridad, y que conecta directamente con la discusión sobre admisibilidad probatoria desarrollada en el capítulo anterior de este libro: mientras el uso de identificación biométrica remota en tiempo real por parte de las fuerzas de seguridad se encuentra sujeto, en la Unión Europea, a un régimen de prohibición general con excepciones estrictamente delimitadas y autorización judicial previa, en la mayoría de los países de la región iberoamericana ese mismo uso no enfrenta, hasta la fecha, ninguna restricción legal específica equivalente, quedando sujeto únicamente a los marcos generales de protección de datos personales y de derechos constitucionales, que no siempre contemplan de manera explícita las particularidades de la tecnología biométrica examinada en este libro.

Esta laguna regulatoria no debería interpretarse como una invitación a que las instituciones forenses de la región adopten estas tecnologías sin restricción alguna, aprovechando la ausencia de un marco legal específico, sino como una oportunidad para que la región, al legislar sobre esta materia, aprenda directamente de la experiencia comparada más avanzada, en lugar de reproducir el patrón, ya documentado en capítulos anteriores de este libro a propósito de la reforma procesal penal, de adoptar tecnologías con años de anticipación respecto de los marcos legales llamados a regularlas.

Un factor adicional que merece consideración específica en el contexto regulatorio latinoamericano es la relación entre la nueva legislación sobre inteligencia artificial y los marcos de protección de datos personales ya existentes en la mayoría de los países de la región, muchos de ellos inspirados, con mayor o menor fidelidad, en el Reglamento General de Protección de Datos de la Unión Europea. Estos marcos de protección de datos, aunque no fueron diseñados originalmente pensando en la inteligencia artificial forense, ya contienen principios directamente aplicables, como la exigencia de una base legal explícita para el tratamiento de datos biométricos, categorizados en la mayoría de estas leyes como datos sensibles que requieren un nivel de protección reforzado, y el derecho de las personas a solicitar información sobre el tratamiento automatizado de sus datos personales. Una estrategia regulatoria pragmática para los países de la región, mientras se desarrolla una legislación específica sobre inteligencia artificial equivalente al reglamento europeo, consistiría en interpretar y aplicar de manera rigurosa estos marcos de protección de datos ya existentes a los usos forenses de inteligencia artificial examinados en este libro, en lugar de esperar pasivamente a que una ley completamente nueva resuelva un vacío que, en gran medida, los marcos actuales ya podrían mitigar si se aplicaran con el rigor adecuado.

Esta estrategia de aplicación rigurosa de los marcos existentes exige, sin embargo, un fortalecimiento institucional de las autoridades de protección de datos personales de la región, muchas de las cuales operan hoy con recursos humanos y técnicos considerablemente más limitados que sus contrapartes europeas, y con una experiencia todavía incipiente en la supervisión específica de sistemas de inteligencia artificial forense. Sin ese fortalecimiento institucional paralelo, la mera existencia de un marco legal de protección de datos aplicable en teoría a estas tecnologías corre el riesgo de convertirse en una garantía puramente formal, incapaz de traducirse en una supervisión efectiva de la práctica cotidiana de las instituciones forenses, un riesgo que este capítulo ya identificó, en la sección anterior, a propósito de la brecha entre la letra y la aplicación efectiva del

propio reglamento europeo.

9.6. Hacia un marco ético transversal para la criminalística con inteligencia artificial

A partir de la discusión desarrollada en este capítulo y en los ocho capítulos anteriores de este libro, es posible sistematizar un conjunto de principios éticos transversales que deberían orientar la incorporación de cualquier sistema de inteligencia artificial a la práctica forense, independientemente de la disciplina específica de que se trate. El primero es el principio de validación previa, desarrollado extensamente en el primer capítulo y retomado en cada capítulo subsiguiente: ningún sistema debería adoptarse sin evidencia empírica independiente de su exactitud y de su desempeño diferencial entre subgrupos relevantes. El segundo es el principio de supervisión humana significativa, que exige que la responsabilidad final de cualquier conclusión forense siga recayendo en un experto humano capacitado, y no en el sistema algorítmico que lo asiste.

Conviene subrayar que estos cinco principios no operan de manera independiente entre sí, sino que se refuerzan mutuamente de un modo que conviene hacer explícito. La validación previa sin supervisión humana significativa corre el riesgo de convertirse en una formalidad vacía, un certificado de calidad que nadie vuelve a verificar una vez obtenido. La supervisión humana sin proporcionalidad puede terminar legitimando el uso rutinario de tecnologías extraordinariamente intrusivas para delitos menores, simplemente porque un ser humano firma el resultado final. La proporcionalidad sin transparencia institucional carece de cualquier mecanismo externo que permita verificar si efectivamente se respeta en la práctica. Y la transparencia sin equidad explícita puede limitarse a informar cifras de uso sin jamás abordar la pregunta normativa de fondo sobre qué noción de justicia esas cifras deberían perseguir. Los cinco principios, en definitiva, constituyen un sistema integrado, y la ausencia de cualquiera de ellos debilita sensiblemente la eficacia protectora de los cuatro restantes.

El tercer principio es el de proporcionalidad, según el cual el grado de intrusión que una tecnología representa sobre los derechos fundamentales de las personas debe ser proporcional a la gravedad del delito investigado y a la necesidad demostrada de esa tecnología específica frente a alternativas menos invasivas, un principio que el marco europeo examinado en este capítulo incorpora de manera explícita al exigir autorización judicial previa para los usos más intrusivos de biometría remota. El cuarto principio es el de transparencia institucional, que exige que las instituciones forenses informen de manera periódica y auditable sobre el tipo y la frecuencia de uso de estas tecnologías, sin comprometer evidentemente la reserva de investigaciones específicas en curso. El quinto y último principio es el de equidad explícita, que exige reconocer, en línea con el dilema revelado por el caso COMPAS que abrió este capítulo, que ninguna solución puramente técnica puede resolver por sí sola las tensiones normativas entre distintas nociones de equidad, y que esas tensiones exigen deliberación pública, jurídica y democrática, y no una decisión delegada silenciosamente a un equipo de desarrollo tecnológico.

Estos cinco principios, formulados de manera abstracta, adquieren su verdadero valor práctico cuando se traducen en preguntas concretas que cualquier institución forense debería poder responder antes de adoptar un nuevo sistema de inteligencia artificial: ¿existe evidencia independiente de validación específica para la población que este sistema atenderá efectivamente? ¿Queda claramente establecido que la decisión final permanece en manos de un experto humano responsable? ¿La intrusión que este sistema representa resulta proporcional a la gravedad de los delitos para cuya investigación se destina? ¿Existe un mecanismo institucional que permita a

la ciudadanía conocer, de manera agregada, cómo y con qué frecuencia se utiliza este sistema? ¿Se ha deliberado explícitamente, y no delegado tácitamente, sobre qué noción de equidad debe priorizar este sistema quien lo adopta? Una institución capaz de responder afirmativamente a las cinco preguntas cuenta con una base ética razonable para adoptar la tecnología en cuestión; una institución que no pueda responder a alguna de ellas debería, como mínimo, suspender la adopción hasta poder hacerlo, sin importar cuán prometedora resulte la tecnología desde una perspectiva puramente técnica.

9.7. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la confusión entre validez técnica y justicia algorítmica, que puede llevar a considerar éticamente aceptable un sistema simplemente por ser estadísticamente exacto, la delegación implícita de decisiones normativas sobre equidad a equipos de desarrollo tecnológico sin deliberación pública, la adopción de tecnologías de vigilancia biométrica en la región iberoamericana sin las salvaguardas de autorización judicial previa y evaluación de impacto que el marco europeo ha comenzado a exigir, y la persistencia de una brecha regulatoria considerable entre la velocidad de adopción tecnológica y la velocidad de desarrollo de marcos legales específicos en la región.

Un riesgo adicional, de particular relevancia para el cierre de este bloque de capítulos dedicados a los desafíos de la inteligencia artificial forense, es el riesgo de fatiga regulatoria: la acumulación, a lo largo de los capítulos siete, ocho y nueve de este libro, de un número considerable de exigencias de validación, admisibilidad y gobernanza ética puede generar, en instituciones forenses con recursos limitados, la percepción de que el cumplimiento simultáneo de todas esas exigencias resulta simplemente inalcanzable, lo que paradójicamente podría desincentivar cualquier esfuerzo de cumplimiento parcial. Frente a este riesgo, conviene recordar que ninguna institución debe satisfacer de manera perfecta y simultánea la totalidad de las exigencias examinadas en estos capítulos antes de dar cualquier paso, sino que puede avanzar de manera progresiva y priorizada, comenzando por las aplicaciones de mayor riesgo para los derechos fundamentales, como la biometría remota y el análisis predictivo basado en personas, y extendiendo gradualmente esas mismas exigencias de validación y gobernanza al resto de las herramientas algorítmicas que vaya incorporando.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan exigir que cualquier decisión sobre qué métrica de equidad priorizar en un sistema de evaluación de riesgo se someta a deliberación pública y jurídica explícita, y no a una decisión técnica implícita, adoptar el modelo de autorización judicial previa y evaluación de impacto sobre derechos fundamentales para cualquier uso de biometría remota en investigación criminal, impulsar activamente el desarrollo de marcos regulatorios específicos para la inteligencia artificial forense en los países de la región que todavía carecen de ellos, y mantener registros institucionales transparentes y auditables sobre el uso de estas tecnologías, como mecanismo de rendición de cuentas ante la ciudadanía.

A estas prácticas conviene sumar una recomendación de cierre para este bloque completo de capítulos dedicados a los desafíos de la inteligencia artificial forense. Ninguna institución debería tratar la ética, el sesgo y los derechos humanos como una consideración posterior, revisada únicamente después de que una tecnología ya haya sido adoptada y desplegada, sino como una dimensión que debe evaluarse en paralelo, y con el mismo rigor, que la validación técnica examinada en el primer

capítulo de este libro y la admisibilidad probatoria examinada en el capítulo anterior. Las tres dimensiones, validez técnica, admisibilidad jurídica y justicia algorítmica, no son etapas sucesivas de un mismo proceso de evaluación, sino tres lentes complementarias que deben aplicarse de manera simultánea a cualquier sistema de inteligencia artificial antes de que ese sistema llegue a influir, de cualquier manera, en la libertad de una persona sometida a un proceso penal.

CAPÍTULO 10

El laboratorio 4.0 y el perito del futuro: gestión del cambio institucional, nuevas competencias y prospectiva 2035-2050

10.1. Caso de apertura: lo que reveló el escándalo de Colorado sobre la gestión institucional del cambio

En 2023, la Oficina de Investigación de Colorado descubrió que Yvonne Missy Woods, una científica forense con treinta años de antigüedad en su laboratorio de ADN, había manipulado resultados genéticos en cientos de casos a lo largo de su carrera, una conducta que derivó en más de cien cargos penales en su contra. La investigación posterior no atribuyó el episodio únicamente a la conducta individual de una examinadora, sino que reveló un problema institucional más profundo: el laboratorio operaba, desde hacía años, con un déficit crónico de personal que generaba una acumulación de casos pendientes en prácticamente todas sus disciplinas, un entorno de presión sostenida en el que las irregularidades de una sola analista permanecieron sin detectarse durante un período extraordinariamente prolongado (Vestal, 2025).

Este episodio, ocurrido en uno de los sistemas forenses mejor financiados y más tecnológicamente avanzados del mundo, ofrece una lección que ningún capítulo anterior de este libro podría transmitir con la misma fuerza: la incorporación de inteligencia artificial a la criminalística, con todo el potencial genuino que los nueve capítulos anteriores han documentado, no puede compensar por sí sola las carencias estructurales de gestión institucional, de capital humano y de supervisión organizacional que afectan a buena parte de los laboratorios forenses del mundo, incluidos aquellos que cuentan con recursos considerablemente superiores a los disponibles en la mayoría de los países de la región iberoamericana. Una evaluación de necesidades encargada por el propio Congreso de Estados Unidos había estimado, ya en 2019, que los laboratorios forenses públicos del país necesitarían más de novecientas posiciones adicionales de tiempo completo únicamente para procesar de manera eficiente su carga de trabajo entonces existente (Office of Justice Programs, 2019), un déficit que la irrupción de nuevas tecnologías, lejos de reducir, tiende a intensificar, al exigir simultáneamente personal capacitado en las disciplinas forenses tradicionales y en las competencias digitales y algorítmicas que este libro ha descrito a lo largo de sus capítulos.

Conviene subrayar un elemento adicional del caso de Colorado que resulta particularmente instructivo para este capítulo final. Las irregularidades de la examinadora permanecieron sin detectarse durante un período que se extendió a lo largo de buena parte de su carrera de tres décadas, pese a que el laboratorio contaba, en principio, con los mismos protocolos de control de calidad y de revisión por pares que este libro ha defendido repetidamente como salvaguarda frente al error humano y al sesgo cognitivo, descritos en detalle en el primer capítulo de esta obra. La pregunta que este episodio plantea no es, entonces, si esos protocolos existían sobre el papel, sino si una institución sometida a una presión de volumen de trabajo sostenida durante años cuenta efectivamente con el tiempo y los recursos humanos necesarios para aplicarlos con el rigor que exigen, o si esa misma presión termina, en la práctica cotidiana, erosionando silenciosamente unos controles de calidad que en teoría deberían proteger tanto a los acusados como a las víctimas de cualquier proceso penal. Esta es, en definitiva, la pregunta institucional de fondo que este capítulo final busca responder.

Este capítulo final de contenido cierra el recorrido de este libro examinando, precisamente, la dimensión institucional y humana de la transformación descrita en los nueve capítulos anteriores: cómo gestionar el cambio organizacional que la incorporación de inteligencia artificial exige, qué nuevas competencias debe desarrollar el perito forense del futuro, y qué escenarios resultan plausibles para la criminalística en las próximas décadas.

Conviene explicar por qué este libro reserva precisamente para su último capítulo de contenido la discusión sobre gestión institucional y formación pericial, en lugar de abrirlo con ella. A lo largo de los nueve capítulos anteriores, este libro ha construido de manera progresiva un argumento completo

sobre qué significa incorporar la inteligencia artificial a la criminalística de manera responsable: comprender sus fundamentos técnicos, validar su desempeño en cada disciplina específica, garantizar su admisibilidad jurídica y su compatibilidad con los derechos fundamentales. Solo después de recorrer ese argumento completo resulta posible formular, con la precisión que este capítulo final se propone, qué tipo de institución y qué tipo de profesional se necesitan para llevarlo efectivamente a la práctica, una pregunta que carecería de sentido si se hubiera formulado antes de establecer, capítulo por capítulo, qué es exactamente lo que esa institución y ese profesional deben ser capaces de hacer.

10.2. El laboratorio 4.0 como transformación organizacional, no solo tecnológica

La expresión laboratorio 4.0, tomada por analogía del concepto de industria 4.0 que describe la integración de sistemas digitales, datos masivos e inteligencia artificial en los procesos productivos, se utiliza cada vez con mayor frecuencia en la literatura forense para describir la transformación digital de los laboratorios criminalísticos. Una síntesis reciente de las discusiones sostenidas en el vigésimo tercer congreso trienal de la Asociación Internacional de Ciencias Forenses advierte, sin embargo, que ningún laboratorio individual cuenta con la capacidad de mantenerse al día, por sí solo, con la velocidad de desarrollo tecnológico descrita a lo largo de este libro, lo que exige fomentar y facilitar una cooperación mucho más amplia entre instituciones de la que ha sido habitual hasta ahora en la criminalística tradicional (Klasén et al., 2024).

Esa misma síntesis concluye con una observación que conecta directamente con el caso de apertura de este capítulo: la incorporación exitosa de inteligencia artificial a un laboratorio forense depende, de manera decisiva, de la valentía y la capacidad de su liderazgo institucional no solo para proponer estrategias de digitalización, sino para efectivamente implementarlas, superando la resistencia organizacional al cambio, la escasez de personal capacitado y la limitación de recursos que caracteriza a la mayoría de los laboratorios forenses del mundo (Klasén et al., 2024). Esta observación desplaza el centro de gravedad de la discusión sobre inteligencia artificial forense, que los capítulos anteriores de este libro concentraron predominantemente en cuestiones de validez técnica, admisibilidad jurídica y ética algorítmica, hacia una dimensión igualmente decisiva pero frecuentemente subestimada: la capacidad de gestión organizacional de las instituciones llamadas a implementar todo lo que este libro ha descrito.

Conviene precisar que hablar de laboratorio 4.0 no implica sugerir que la transformación digital de la criminalística deba producirse de manera uniforme e idéntica en todos los contextos institucionales, una precisión particularmente relevante para la región iberoamericana. Un laboratorio central de una capital nacional, con acceso a financiamiento internacional y a personal formado en universidades con programas de posgrado especializados, enfrenta un punto de partida radicalmente distinto al de una delegación regional con escasos recursos y alta rotación de personal, un contraste que este libro ya documentó en su primer capítulo a propósito de la desigualdad interna que caracteriza a buena parte de los sistemas forenses de la región. La transformación hacia el laboratorio 4.0 debería, en consecuencia, entenderse no como un destino único e idéntico para toda institución forense, sino como una trayectoria de mejora continua cuyo punto de partida y cuyo ritmo de avance dependen, necesariamente, de las condiciones institucionales concretas de cada laboratorio, sin que eso justifique renunciar a la aspiración de mejora que este libro ha defendido en cada uno de sus capítulos.

Esta heterogeneidad de puntos de partida sugiere, además, que la métrica más adecuada para evaluar el progreso de un laboratorio hacia el modelo 4.0 no debería ser una comparación absoluta con los laboratorios mejor equipados del mundo, una comparación que condenaría a la inmensa

mayoría de las instituciones de la región a una sensación perpetua de insuficiencia, sino una medición relativa de la trayectoria de mejora propia de cada institución a lo largo del tiempo, evaluada contra su propio punto de partida y contra los recursos efectivamente disponibles en su contexto específico. Este cambio de métrica, de la comparación absoluta a la trayectoria relativa, no es un mero matiz retórico, sino una decisión de gestión con consecuencias prácticas directas sobre la motivación del personal y sobre la viabilidad política de sostener, en el tiempo, cualquier proceso de transformación institucional tan exigente como el que este libro ha descrito a lo largo de sus diez capítulos.

10.3. Gestión del cambio institucional: lecciones de la transformación digital en laboratorios forenses

Un marco de gobernanza propuesto recientemente por especialistas convocados por INTERPOL para la incorporación de inteligencia artificial y grandes modelos de lenguaje a la informática forense digital ofrece un conjunto de recomendaciones directamente aplicables a la gestión del cambio institucional examinada en este capítulo. Ese marco identifica como responsabilidades centrales del liderazgo de un laboratorio la gestión del capital humano y del cambio organizacional, que incluyen el desarrollo de repositorios de procedimientos operativos estandarizados, calendarios sistemáticos de mantenimiento y actualización de las herramientas algorítmicas empleadas, y programas de incorporación, formación y reentrenamiento del personal alineados con los ciclos de acreditación institucional (INTERPOL, 2025).

Este mismo marco de gobernanza detalla, además, los requisitos específicos de control de cambios que deberían regir cualquier flujo de trabajo forense asistido por inteligencia artificial o por grandes modelos de lenguaje, entre ellos la documentación de la procedencia de los datos de entrenamiento utilizados por cada herramienta, un mecanismo de control de cambios formal para cualquier actualización de versión, la monitorización continua de lo que la literatura técnica denomina desviación del modelo, es decir, el deterioro progresivo del desempeño de un sistema a medida que las condiciones reales de operación se alejan de las condiciones bajo las cuales fue originalmente entrenado y validado, y un sistema de registro auditable que documente con precisión qué proporción de cada conclusión pericial corresponde al juicio del examinador y qué proporción corresponde a la sugerencia del sistema algorítmico (INTERPOL, 2025). Esta última exigencia retoma, en clave de gestión operativa cotidiana, la misma discusión sobre cadena de custodia algorítmica que el octavo capítulo de este libro desarrolló desde una perspectiva jurídica y probatoria, confirmando que ambas dimensiones, la institucional y la legal, convergen en exigencias prácticas muy similares.

Este mismo marco recomienda, además, la designación de personal específicamente encargado de liderar la adopción de inteligencia artificial dentro de cada unidad operativa, trabajando en paralelo con una inversión sostenida en el desarrollo de las competencias del resto del personal, de manera que la capacidad de analizar críticamente los resultados de un sistema algorítmico y de comunicar su incertidumbre de manera adecuada, un tema que atraviesa buena parte de los capítulos anteriores de este libro, no quede concentrada en un pequeño grupo de especialistas técnicos, sino que se distribuya de manera razonable a lo largo de toda la organización (INTERPOL, 2025). Esta recomendación resulta directamente aplicable al escenario descrito en el caso de apertura de este capítulo: un sistema de supervisión distribuido, en el que múltiples miembros del personal cuentan con la capacitación necesaria para identificar anomalías en el trabajo de sus colegas, ofrece una salvaguarda institucional considerablemente más robusta que la concentración de esa responsabilidad en un único examinador, por experimentado que sea, cuyo trabajo termina sin supervisión efectiva durante años.

Investigaciones etnográficas sobre procesos de transformación digital en laboratorios de diagnóstico clínico, un dominio institucional con desafíos organizacionales estrechamente comparables a los del laboratorio forense, documentan que los procesos de digitalización exitosos comparten una característica común: en lugar de imponer el cambio tecnológico de manera abrupta y centralizada, el liderazgo institucional dedica tiempo explícito a atender las necesidades del personal directamente afectado por la transformación, mediante formación continua y disponibilidad constante para resolver dificultades técnicas durante el período de transición, un enfoque que permite que la hibridez entre los procesos antiguos y los nuevos, inevitable durante cualquier proceso de cambio organizacional profundo, se gestione de manera ordenada en lugar de generar resistencia y confusión prolongadas entre el personal.

Esta misma investigación etnográfica documenta un hallazgo adicional que merece destacarse por su relevancia directa para el argumento de este capítulo: la transición hacia sistemas completamente digitales rara vez ocurre de manera instantánea, sino que atraviesa un período prolongado de convivencia entre procesos antiguos en papel y procesos nuevos digitales, una hibridez que el personal de las instituciones estudiadas terminó considerando no un fracaso o una demora indeseable del proceso de transformación, sino una oportunidad genuina para tender puentes entre generaciones distintas de profesionales, permitiendo que el personal con mayor experiencia en los métodos tradicionales transfiriera su conocimiento acumulado a las nuevas generaciones, mientras estas últimas facilitaban, a su vez, la adaptación del personal más experimentado a las nuevas herramientas digitales. Trasladada al contexto de los laboratorios forenses examinados en este libro, esta observación sugiere que un período de hibridez gestionado deliberadamente, en lugar de una transición forzosamente abrupta, podría constituir una estrategia de gestión del cambio más efectiva que la sustitución inmediata y completa de los métodos tradicionales por los sistemas de inteligencia artificial descritos a lo largo de este libro.

10.4. La crisis de capital humano: escasez, acumulación de casos pendientes y el papel de la inteligencia artificial

El déficit de personal documentado en el caso de apertura de este capítulo no constituye un problema exclusivo de Estados Unidos, sino una tensión estructural que afecta, con distinta intensidad, a laboratorios forenses de todo el mundo, y que adquiere una relevancia particular para la región iberoamericana, donde las restricciones presupuestarias examinadas en capítulos anteriores de este libro suelen ser considerablemente más severas que las de los sistemas forenses de países de mayores ingresos. Esta tensión plantea una pregunta que atraviesa buena parte del argumento de este capítulo: ¿puede la inteligencia artificial aliviar genuinamente esta crisis de capital humano, o corre el riesgo de intensificarla, al exigir competencias adicionales que el personal existente no posee, sin que su incorporación reduzca de manera proporcional la carga de trabajo que ese mismo personal debe sostener?

Conviene situar esta tensión en el contexto más amplio de la profesión forense como carrera, un ángulo que ninguna sección anterior de este libro ha explorado directamente. La escasez de personal documentada en el caso de apertura de este capítulo no responde únicamente a limitaciones presupuestarias, sino también a un desafío de atracción y retención de talento que la propia naturaleza del trabajo forense agrava: se trata de una profesión que exige una formación científica considerablemente exigente, que expone a quienes la ejercen a un volumen de trabajo creciente y a menudo emocionalmente demandante, particularmente en disciplinas como la medicina forense o la investigación de fosas clandestinas examinada en el tercer capítulo de este libro, y que en numerosos países de la región ofrece una remuneración considerablemente inferior a la que un profesional con formación científica equivalente podría obtener en el sector privado

o en la industria tecnológica. La incorporación de inteligencia artificial, lejos de resolver este desafío de atracción de talento por sí sola, podría incluso agravarlo en el corto plazo, al generar la percepción errónea entre estudiantes de ciencias de que la automatización reducirá la demanda futura de peritos humanos, una percepción que este libro ha refutado capítulo tras capítulo, pero que las instituciones de la región deberían abordar de manera proactiva en sus estrategias de comunicación y reclutamiento.

La evidencia recogida a lo largo de este libro sugiere una respuesta matizada, ni excesivamente optimista ni excesivamente pesimista. En tareas específicas y bien delimitadas, como la detección automática de regiones de interés en imágenes balísticas descrita en el sexto capítulo o la clasificación preliminar de contenido de interés forense en escenas extensas descrita en el quinto capítulo, la inteligencia artificial ha demostrado una capacidad genuina de reducir el tiempo que un examinador humano debe dedicar a tareas repetitivas de bajo valor analítico, liberando ese tiempo para las etapas de interpretación y verificación que, según ha insistido este libro capítulo tras capítulo, deben permanecer bajo responsabilidad humana. Sin embargo, esa ganancia de eficiencia no se traduce automáticamente en una reducción proporcional de la acumulación de casos pendientes, si la implementación de la nueva tecnología no viene acompañada de la inversión formativa necesaria para que el personal existente pueda efectivamente aprovecharla, un riesgo que el propio marco de gobernanza de INTERPOL examinado en la sección anterior identifica como una de las principales causas de fracaso en los procesos de digitalización forense.

Conviene señalar, además, que la acumulación de casos pendientes no afecta de manera uniforme a todas las disciplinas forenses examinadas a lo largo de este libro, sino que se concentra de manera particularmente severa en aquellas que dependen de análisis de laboratorio intensivos en tiempo, como la genética forense descrita en el cuarto capítulo, mientras que otras disciplinas, como determinados componentes de la informática forense digital, han logrado avances de automatización más sustanciales en los últimos años. Esta heterogeneidad exige que cualquier estrategia institucional de reducción de rezagos, apoyada o no en inteligencia artificial, identifique con precisión en qué disciplina específica se concentra el cuello de botella de cada laboratorio particular, en lugar de aplicar una misma solución tecnológica genérica a la totalidad de sus operaciones, un principio de diagnóstico previo a la intervención que resulta coherente con la lógica de validación específica por aplicación que este libro ha defendido desde su primer capítulo.

10.5. El perito del futuro: nuevas competencias y alfabetización algorítmica

A lo largo de los nueve capítulos anteriores de este libro ha aparecido, de manera recurrente, una misma recomendación formativa: ningún perito debería operar o evaluar críticamente un sistema de inteligencia artificial sin haber alcanzado antes un nivel mínimo de alfabetización técnica sobre sus fundamentos, sus limitaciones y sus condiciones de validación. Este capítulo sistematiza esa recomendación en un perfil de competencias concreto para el perito forense de la próxima década, que complementa, sin sustituir, el dominio disciplinario tradicional que cualquier especialista forense debe seguir poseyendo.

Ese perfil de competencias debería incluir, como mínimo, cuatro dimensiones desarrolladas a lo largo de este libro. La primera es la comprensión conceptual de los fundamentos de la inteligencia artificial presentados en el segundo capítulo, suficiente para distinguir entre aprendizaje automático y aprendizaje profundo, para comprender la diferencia entre datos de entrenamiento y datos de prueba, y para identificar los riesgos característicos de cada tipo de arquitectura algorítmica. La segunda es la alfabetización estadística necesaria para interpretar correctamente una razón de verosimilitud, una tasa de error condicionada al contexto del caso, o una métrica de desempeño

diferencial entre subgrupos demográficos, competencias desarrolladas en los capítulos tercero, cuarto y sexto de este libro. La tercera es la capacidad de comunicación de la incertidumbre, indispensable para explicar ante un tribunal, en lenguaje accesible y sin simplificaciones engañosas, el margen de error y las limitaciones de cualquier conclusión asistida por inteligencia artificial, una competencia central en la discusión sobre admisibilidad probatoria del octavo capítulo. La cuarta es la sensibilidad ética y de derechos humanos necesaria para identificar cuándo una herramienta técnicamente válida puede, pese a ello, generar consecuencias discriminatorias o desproporcionadas, la dimensión desarrollada en el noveno capítulo de este libro.

Conviene precisar que estas cuatro competencias no exigen que cada perito forense se convierta en un especialista en ciencia de datos o en ingeniería de sistemas, una expectativa tan poco realista como innecesaria. Lo que exigen es un nivel de alfabetización suficiente para formular las preguntas correctas frente a cualquier herramienta algorítmica, de manera análoga a como un médico clínico no necesita ser farmacólogo molecular para prescribir un medicamento de manera responsable, pero sí necesita comprender sus mecanismos generales de acción, sus contraindicaciones y sus márgenes de seguridad. Esta analogía, lejos de ser meramente ilustrativa, sugiere un modelo formativo concreto: así como la farmacología clínica constituye hoy una asignatura obligatoria en cualquier programa de medicina, sin que ello convierta a cada médico graduado en farmacólogo, la alfabetización algorítmica forense debería constituir un componente obligatorio y transversal de cualquier programa de formación en criminalística, integrado a lo largo de la currícula y no relegado a un curso optativo aislado que solo cursan quienes ya sienten una inclinación previa hacia lo tecnológico.

Investigaciones sobre la integración de inteligencia artificial en la educación forense y en ciencias afines documentan que herramientas como los sistemas de tutoría inteligente, la evaluación automatizada y los entornos de simulación virtual ofrecen un potencial considerable para acelerar el desarrollo de estas competencias, al permitir a los estudiantes practicar la interpretación de resultados algorítmicos y la comunicación de incertidumbre en escenarios simulados antes de enfrentarse a las consecuencias reales de un caso judicial (Rajarajeshwari y Architha, 2025).

Un componente adicional de este perfil de competencias, todavía poco discutido en la literatura especializada pero que se desprende con claridad de la argumentación desarrollada a lo largo de este libro, es la capacidad de colaboración interdisciplinaria. Ningún perito individual, por completa que sea su formación en las cuatro dimensiones descritas en este capítulo, puede dominar en profundidad la totalidad de las disciplinas forenses examinadas en este libro, desde la genética hasta la informática forense digital pasando por la balística y la documentoscopia. La criminalística del futuro exige, en consecuencia, no solo peritos individualmente mejor formados, sino equipos capaces de trabajar de manera coordinada entre especialidades distintas, apoyados en plataformas compartidas que faciliten precisamente ese tipo de colaboración, una capacidad organizacional que conecta directamente con la discusión sobre gestión del cambio institucional desarrollada en la primera mitad de este capítulo, y que confirma, una vez más, que la dimensión técnica y la dimensión institucional de la transformación descrita en este libro resultan, en última instancia, inseparables.

Algunas instituciones académicas han comenzado ya a rediseñar sus programas de formación forense pensando explícitamente en este nuevo perfil de competencias: la Universidad Hallym de Corea del Sur estableció en 2023 un departamento específicamente dedicado a la ciencia y tecnología de la información forense, con un plan de estudios que integra materias tradicionales de investigación policial con cursos avanzados de inteligencia artificial aplicada, seguridad de la información y política de datos, formando profesionales capaces de desenvolverse tanto en la

investigación criminal como en el amplio abanico de industrias digitales que hoy generan evidencia con potencial valor forense.

Para la región iberoamericana, la creación de programas específicamente diseñados según este nuevo perfil de competencias, en lugar de la simple incorporación de un curso aislado de inteligencia artificial dentro de programas de criminalística ya existentes, representa una oportunidad de posicionamiento académico considerable, particularmente si esos programas incorporan desde su diseño inicial los desafíos específicos que este libro ha documentado para la región: la escasa representatividad de las poblaciones latinoamericanas en los conjuntos de entrenamiento de numerosas herramientas comerciales, discutida en los capítulos tercero y cuarto, la brecha de infraestructura entre laboratorios centrales y regionales, discutida en el capítulo quinto, y la ausencia todavía generalizada de marcos regulatorios específicos, discutida en los capítulos octavo y noveno. Un programa de formación que prepare a sus egresados no solo para operar herramientas de inteligencia artificial desarrolladas en otras regiones del mundo, sino para evaluarlas críticamente, adaptarlas y, en el mejor de los escenarios, contribuir al desarrollo de herramientas propias validadas específicamente para las poblaciones y los contextos institucionales de la región, constituiría una contribución genuina al cierre de la brecha tecnológica que este libro ha documentado a lo largo de sus diez capítulos.

10.6. Prospectiva 2035-2050: escenarios posibles para la criminalística

Cualquier ejercicio de prospectiva a un plazo de una o dos décadas exige un grado considerable de humildad epistémica, particularmente en un campo cuya velocidad de cambio tecnológico este libro ha documentado repetidamente. Sin dejar de reconocer esa incertidumbre, es posible trazar, a partir de la evidencia acumulada a lo largo de los nueve capítulos anteriores, tres escenarios amplios y no necesariamente excluyentes que podrían caracterizar a la criminalística de las próximas dos décadas.

El primer escenario, que podría denominarse de integración madura, proyecta una consolidación progresiva de los estándares de validación, admisibilidad y gobernanza ética examinados en los capítulos séptimo, octavo y noveno de este libro, de manera que hacia 2035 la mayoría de los sistemas de justicia, incluidos los de la región iberoamericana, cuenten con marcos legislativos específicos para la prueba algorítmica, mecanismos de acreditación dinámica para herramientas de inteligencia artificial forense, y una generación de peritos formados desde el inicio de su carrera en las competencias descritas en la sección anterior. Este escenario, el más favorable de los tres, no es automático ni inevitable, sino que depende de decisiones de política pública e inversión institucional que este libro ha intentado justificar a lo largo de sus diez capítulos.

En este escenario de integración madura, cabría esperar que hacia 2040 y 2050 la criminalística haya desarrollado un cuerpo de investigación empírica propio, comparable en rigor al que hoy sostiene a la genética forense clásica, para la mayoría de las disciplinas asistidas por inteligencia artificial examinadas en este libro, de manera que la crisis de validez descrita en el primer capítulo se convierta definitivamente en un episodio histórico superado, y no en un problema recurrente que cada nueva generación de tecnología forense deba enfrentar desde cero. En ese horizonte, es razonable anticipar además una integración considerablemente mayor entre las disciplinas forenses tradicionalmente separadas que este libro ha examinado capítulo por capítulo, de manera que un mismo caso de investigación pueda combinar, dentro de una plataforma unificada de análisis, evidencia balística, genética, digital y de reconstrucción de escena, procesada de manera coordinada por sistemas de inteligencia artificial interoperables entre sí, en lugar del enfoque compartimentado por disciplina que todavía predomina en la práctica forense actual.

El segundo escenario, de fragmentación regulatoria persistente, proyecta la continuidad de la brecha ya documentada en el noveno capítulo de este libro entre jurisdicciones con marcos regulatorios avanzados, como la Unión Europea, y jurisdicciones que continúan adoptando tecnología forense sin marcos legales equivalentes, un escenario en el cual la validez científica y la protección de derechos fundamentales de una persona sometida a un proceso penal dependerían, de manera creciente, del país específico donde ese proceso se desarrolle, profundizando desigualdades ya existentes entre sistemas de justicia de distintos niveles de desarrollo institucional. El tercer escenario, de sustitución acelerada sin salvaguardas suficientes, proyecta una adopción masiva y poco regulada de sistemas de inteligencia artificial cada vez más autónomos, en la que la supervisión humana significativa defendida en cada capítulo de este libro se erosiona progresivamente por razones de costo y de volumen de casos, reproduciendo a gran escala los riesgos de sesgo, opacidad y error que el primer capítulo de este libro documentó para la criminalística tradicional, ahora amplificados por la velocidad y el alcance de los sistemas algorítmicos.

Resulta importante subrayar que estos tres escenarios no constituyen categorías mutuamente excluyentes aplicables de manera uniforme a un país entero, sino que es perfectamente posible, y quizás incluso probable, que distintas disciplinas forenses dentro de un mismo sistema de justicia evolucionen según trayectorias distintas, más cercanas a la integración madura en aquellas disciplinas donde la comunidad científica internacional ha invertido mayores recursos de validación, como la genética forense examinada en el cuarto capítulo, y más cercanas a la sustitución acelerada sin salvaguardas suficientes en disciplinas emergentes donde la presión comercial supera todavía a la investigación académica independiente, como algunas de las aplicaciones de inteligencia artificial generativa examinadas en el séptimo capítulo. Anticipar esta posible divergencia interna permite a las instituciones forenses de la región priorizar sus esfuerzos regulatorios y de inversión precisamente en aquellas disciplinas donde el riesgo de derivar hacia el escenario menos favorable resulta más inminente, en lugar de dispersar recursos limitados de manera uniforme entre todas las aplicaciones examinadas en este libro (véase la Figura 10.1).

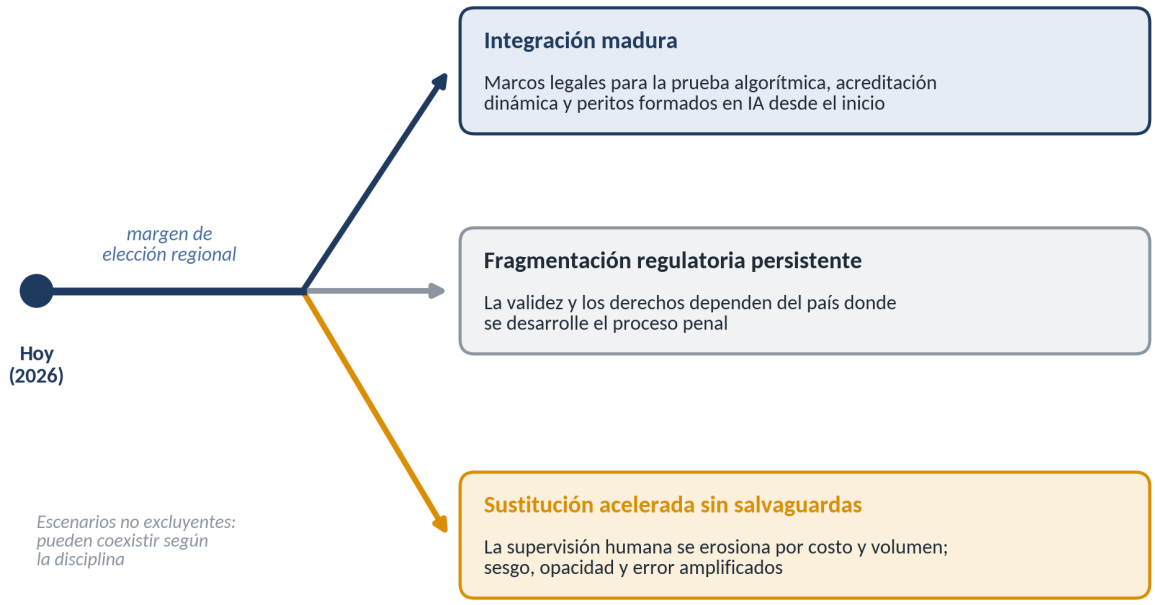


Figura 10.1. Escenarios prospectivos para la criminalística, 2035-2050. Los tres escenarios no son excluyentes: distintas disciplinas de un mismo sistema de justicia pueden seguir trayectorias diferentes. Fuente: elaboración propia.

La región iberoamericana se encuentra hoy, según ha documentado este libro capítulo tras capítulo,

en una posición desde la cual todavía puede optar entre estos tres escenarios, con mayor margen de maniobra que sistemas de justicia que ya han avanzado más profundamente por alguno de estos caminos. Esa es, en definitiva, la invitación central de este libro: que la criminalística de la región construya su propio camino hacia la integración madura, aprendiendo de la experiencia comparada documentada en cada uno de estos diez capítulos, sin necesidad de atravesar primero las mismas crisis de validez, los mismos escándalos institucionales y los mismos episodios de discriminación algorítmica que otras regiones del mundo ya han experimentado.

10.7. Cooperación regional: una vía para acelerar el escenario de integración madura

Un mecanismo institucional que podría acelerar de manera considerable el avance de la región iberoamericana hacia el escenario de integración madura descrito en la sección anterior, y que ningún capítulo previo de este libro ha desarrollado en detalle, es la cooperación regional activa entre los institutos de criminalística de distintos países. A diferencia de Europa, donde el propio proceso de integración regional facilitó, con todas sus imperfecciones, el desarrollo de marcos normativos y técnicos compartidos como los examinados en el noveno capítulo de este libro, la región iberoamericana carece todavía de un mecanismo equivalente específicamente dedicado a la criminalística asistida por inteligencia artificial, más allá de los canales tradicionales de cooperación policial y judicial que atienden, en general, casos individuales de asistencia mutua y no el desarrollo compartido de estándares técnicos.

Una agenda de cooperación regional en esta materia podría avanzar, de manera concreta, en al menos cuatro frentes que sintetizan buena parte de las recomendaciones formuladas a lo largo de este libro. El primero es el desarrollo compartido de conjuntos de datos de entrenamiento y de validación representativos de las poblaciones de la región, que ningún país individual podría reunir con la misma rapidez ni con la misma diversidad demográfica que un esfuerzo coordinado entre varios países, una necesidad señalada de manera reiterada en los capítulos tercero y cuarto de este libro a propósito de la escasa representación de poblaciones latinoamericanas en los modelos comerciales disponibles. El segundo es la armonización progresiva de estándares mínimos de validación y de admisibilidad probatoria para la prueba algorítmica, siguiendo el espíritu, aunque no necesariamente la letra exacta, de la serie ISO 21043 examinada en el octavo capítulo, de manera que un dictamen pericial validado en un país de la región pueda ser reconocido con mayor facilidad por otro, facilitando la cooperación judicial en casos transnacionales de creciente frecuencia, como los examinados a propósito del crimen organizado y de los criptoactivos en el séptimo capítulo de este libro.

El tercer frente de esa agenda de cooperación regional es la creación de capacidades compartidas de formación especializada, mediante programas de posgrado regionales o mediante acuerdos de intercambio entre universidades que permitan a estudiantes y profesionales de distintos países acceder a la formación en alfabetización algorítmica descrita en este capítulo, sin que cada país individual deba desarrollar, de manera aislada y costosa, su propio programa completo desde cero. El cuarto frente, finalmente, es la creación de mecanismos regionales de auditoría algorítmica independiente, análogos a los laboratorios centralizados de auditoría pericial para la defensa pública propuestos en el octavo capítulo de este libro, pero concebidos a escala regional para atender a países cuyo tamaño individual no justificaría, por sí solo, el desarrollo de esa capacidad especializada. Ninguno de estos cuatro frentes exige inversión desproporcionada si se aborda de manera coordinada entre varios países, y cada uno de ellos aceleraría de manera directa el tránsito de la región hacia el escenario de integración madura que este capítulo ha descrito como el más deseable de los tres examinados.

10.8. Riesgos y buenas prácticas

Riesgos identificados en este capítulo

Entre los riesgos centrales que se desprenden de este capítulo se encuentran la expectativa de que la inteligencia artificial pueda compensar, por sí sola, carencias estructurales de capital humano y de gestión institucional, la implementación de tecnología digital sin la inversión formativa paralela necesaria para que el personal existente pueda aprovecharla, la concentración de la capacidad de supervisión algorítmica en un número reducido de especialistas sin distribución razonable de esa responsabilidad a lo largo de la organización, y la persistencia de programas de formación pericial que no incorporan las competencias de alfabetización algorítmica que este capítulo ha sistematizado.

Un riesgo adicional, de cierre para este último capítulo de contenido temático, merece señalarse porque sintetiza una tensión que ha recorrido el libro completo. Existe el riesgo de que la propia extensión y complejidad de las recomendaciones formuladas a lo largo de estos diez capítulos, desde la validación fundacional hasta la cooperación regional pasando por la admisibilidad probatoria y la ética algorítmica, resulte abrumadora para instituciones forenses que operan con recursos modestos y con una carga de trabajo ya considerable, generando el riesgo paradójico de que la magnitud del desafío termine desalentando cualquier acción en absoluto. Frente a ese riesgo final, conviene recordar el mensaje que este libro ha repetido desde su primer capítulo: ninguna institución debe resolver de manera simultánea la totalidad de los desafíos aquí descritos, sino que puede avanzar de manera progresiva, priorizada y adaptada a su contexto específico, con la certeza de que cada paso dado en la dirección correcta, por modesto que sea, representa una mejora genuina respecto de no haberlo dado.

Buenas prácticas recomendadas

Entre las buenas prácticas que se desprenden de este capítulo destacan diseñar cualquier proceso de digitalización forense como una transformación organizacional integral, y no como una simple adquisición de equipamiento, invertir de manera proporcional en formación y reentrenamiento del personal existente antes o en paralelo con la adopción de nuevas herramientas algorítmicas, distribuir la capacidad de supervisión crítica de sistemas de inteligencia artificial a lo largo de toda la organización y no en un grupo reducido de especialistas, e incorporar de manera explícita las cuatro competencias de alfabetización algorítmica sistematizadas en este capítulo a los programas de formación pericial de la región.

A estas prácticas conviene sumar una recomendación final que sintetiza el argumento completo de este capítulo. Ninguna institución forense debería medir el éxito de su transformación hacia el laboratorio 4.0 únicamente en términos de cuántas herramientas de inteligencia artificial ha adoptado, sino en términos de si esa adopción ha venido acompañada de la inversión paralela en personal, formación y gobernanza organizacional que este capítulo ha documentado como condición indispensable para que la tecnología cumpla su promesa. El caso de Colorado que abrió este capítulo, ocurrido en un laboratorio que sin duda contaba con acceso a tecnología forense de vanguardia, demuestra con crudeza que la sofisticación tecnológica sin una gestión institucional igualmente sofisticada no protege a nadie, ni a las personas acusadas, ni a las víctimas, ni a la confianza pública en el sistema de justicia.

Con esta recomendación se cierra el recorrido temático de este libro, que partió, en su primer capítulo, de una crisis de validez científica documentada por los informes NAS y PCAST, y que ha atravesado, a lo largo de diez capítulos, los fundamentos técnicos de la inteligencia artificial, sus aplicaciones específicas en cada disciplina forense, los desafíos digitales emergentes, la

admisibilidad jurídica de la prueba algorítmica, la ética y los derechos humanos, y finalmente la dimensión institucional y humana que este último capítulo ha desarrollado. El epílogo que sigue a continuación ofrece una reflexión final sobre lo que significa, para la criminalística, sostener el compromiso con la verdad científica en un tiempo de realidades cada vez más sintéticas.

EPÍLOGO

La verdad científica en tiempos de realidades sintéticas

Este libro comenzó con un hombre inocente detenido a partir de una huella que no era la suya. Comenzó, más precisamente, con una advertencia de 2009: que buena parte de lo que la criminalística había dado por cierto durante un siglo carecía todavía de la validación científica que exigía llamarlo ciencia. Termina, diez capítulos después, con un laboratorio que colapsó no por falta de tecnología, sino por falta de personas suficientes para vigilar a quienes la usaban. Entre un extremo y otro se despliega el argumento completo de esta obra, y conviene, en este espacio final, decir con toda claridad cuál es.

La tesis de este libro no es que la inteligencia artificial vaya a salvar a la criminalística de sus problemas históricos, ni que vaya a agravarlos de manera irremediable. Es algo más preciso y, por eso mismo, más útil: que la inteligencia artificial reproduce, a una velocidad considerablemente mayor y con una escala considerablemente más amplia, exactamente los mismos problemas que esta disciplina lleva más de un siglo intentando resolver. La necesidad de validar un método antes de confiar en él. La necesidad de reconocer que el juicio humano, entrenado y bien intencionado, es también falible. La necesidad de que la prueba pueda explicarse, cuestionarse y, cuando corresponda, refutarse ante un tribunal. La necesidad de que la ciencia forense sirva a la justicia sin convertirse en un instrumento de discriminación. Ninguna de estas exigencias es nueva. Lo nuevo es la urgencia con la que hoy se presentan, y la sofisticación matemática que ahora se requiere para satisfacerlas.

El título de este epílogo alude a algo que ningún capítulo anterior nombró de manera tan directa, aunque todos lo rozaron: vivimos ya en un tiempo en el que la propia realidad puede fabricarse con una fidelidad que desafía la percepción humana. Un video puede no haber ocurrido nunca. Una voz puede no pertenecer a quien parece pronunciarla. Un documento puede parecer auténtico sin serlo. Esta condición, que el séptimo capítulo de este libro describió con el nombre técnico de contenido sintético, no es un episodio pasajero de la historia tecnológica, sino, según toda la evidencia reunida en estas páginas, un rasgo permanente del paisaje en el que la criminalística deberá operar de ahora en adelante. La pregunta que Edmond Locard nunca tuvo que formularse, porque en su época la propia noción resultaba impensable, es hoy inevitable: ¿qué significa buscar la verdad material de un hecho cuando la propia materialidad de la evidencia puede fabricarse con precisión indistinguible de lo real?

Este libro no ofrece una respuesta definitiva a esa pregunta, porque ninguna respuesta definitiva existe todavía. Ofrece, en cambio, un método para pensarla: el mismo método que Locard inauguró hace más de un siglo, y que los diez capítulos de esta obra han aplicado, uno tras otro, a cada nueva tecnología que la criminalística ha incorporado. Ese método no consiste en maravillarse ante lo nuevo ni en rechazarlo por desconfianza. Consiste en preguntar, cada vez, con la misma disciplina: ¿ha sido validado? ¿Conocemos su tasa de error? ¿Puede explicarse ante quien debe juzgar? ¿Trata a todas las personas con la misma justicia? Y, sobre todo, ¿sigue habiendo, al final de la cadena, un ser humano responsable que responde por la conclusión alcanzada?

Esa última pregunta, la de la responsabilidad humana, es quizás la que mejor resume el espíritu de este libro. En cada uno de los diez capítulos que lo componen, desde la balística hasta la genética, desde el reconocimiento facial hasta el análisis predictivo, la evidencia más sólida disponible ha señalado, una y otra vez, la misma dirección: la inteligencia artificial rinde mejor cuando amplía el criterio humano, y falla, o se vuelve peligrosa, cuando pretende sustituirlo. No porque la tecnología

carezca de mérito, que lo tiene y considerable, sino porque la responsabilidad de decidir sobre la libertad de una persona no es una tarea que pueda delegarse, ni parcial ni totalmente, a un sistema que no puede rendir cuentas de la manera en que un ser humano sí puede hacerlo.

Para la criminalística de Iberoamérica, region que atraviesa este libro como un interlocutor constante y no como una nota al pie, esta tesis tiene una implicancia particular. La región no llega tarde a esta transformación por falta de talento ni de vocación científica, ambos abundantes en sus institutos forenses y sus universidades, sino por falta, todavía, de los marcos institucionales, regulatorios y formativos que este libro ha descrito capítulo tras capítulo. Esa misma tardanza, sin embargo, ofrece una oportunidad que pocas regiones del mundo conservan intacta: la de observar los errores ya cometidos en otros lugares, los escándalos ya documentados, las crisis de validez ya atravesadas, y construir, con ese conocimiento previo, una criminalística que no tenga que volver a aprender por la vía dolorosa lo que este libro ha intentado transmitir por la vía de la lectura.

Quede entonces, como cierre de esta obra, la misma invitación que Locard formuló hace más de un siglo cuando convenció a la policía de Lyon de cederle dos habitaciones para instalar el primer laboratorio forense del mundo: que el indicio material, interrogado con el rigor suficiente, puede decirnos algo verdadero sobre lo que ocurrió. La inteligencia artificial no cambia esa promesa fundacional. La hereda, la amplía y, si se la exige con el mismo rigor que este libro ha defendido en cada una de sus páginas, puede ayudar a cumplirla mejor de lo que ningún examinador humano, por talentoso que sea, podría cumplirla en soledad. Esa es, en definitiva, la criminalística del futuro que este libro ha intentado imaginar: no una que reemplace al perito por la máquina, sino una que, gracias a la máquina, permita por fin que el perito responda, con evidencia y no solo con experiencia, a la pregunta que ha definido a esta disciplina desde su origen. Qué fue lo que realmente ocurrió.

BIBLIOGRAFÍA

Access Now. (2024). Radiografía normativa: ¿Dónde, qué y cómo se está regulando la inteligencia artificial en América Latina? Access Now.

Agarwal, U., Rishiwal, V., Tanwar, S., y Yadav, M. (2024). Blockchain and crypto forensics: Investigating crypto frauds. International Journal of Network Management, 34(2), e2255. https://doi.org/10.1002/nem.2255

Ahmed, A. A., Farhan, K., Jabbar, W. A., Al-Othmani, A., y Abdulrahman, A. G. (2024). IoT forensics: Current perspectives and future directions. Sensors, 24(16), 5210. https://doi.org/10.3390/s24165210

Alduais, A., Al-Khulaidi, M. A., Allegretta, S., y Abdulkhalek, M. M. (2023). Forensic linguistics: A scientometric review. Cogent Arts & Humanities, 10(1). https://doi.org/10.1080/23311983.2023.2214387

Altındağ, A., Öztürk, B., Kazangirler, B. Y., y Pekince, A. (2023). Artificial intelligence application and performance in forensic age estimation with mandibular third molars on radiographs. Balkan Journal of Dental Medicine, 27(3), 181–186. https://doi.org/10.5937/bjdm2303181a

Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016, 23 de mayo). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.

Ataş, İ., Özdemir, C., Ataş, M., y Doğan, Y. (2023). Forensic dental age estimation using modified deep learning neural network. Balkan Journal of Electrical and Computer Engineering. https://doi.org/10.17694/bajece.1351546

Atlam, M., y Wills, G. (2024). Blockchain forensics: A systematic literature review. Electronics, 13(17), 3568.

Baiker-Sørensen, M., Alberink, I., Granell, L. B., van der Ham, L., Mattijssen, E. J. A. T., Smith, E. D., Soons, J., Vergeer, P., y Zheng, X. A. (2023). Automated interpretation of comparison scores for firearm toolmarks on cartridge case primers. Forensic Science International, 353, 111858. https://doi.org/10.1016/j.forsciint.2023.111858

Balbudhe, M. S., Shetty, B. S. K., Arun Kumar, N., Babar, S., Sinha, A., y Gupta, S. (2025). Crime scene reconstruction: A scoping review. Egyptian Journal of Forensic Sciences, 15(1). https://doi.org/10.1186/s41935-025-00495-5

Ballantyne, J., y Butler, J. M. (2015). Toward global forensic DNA standards. En Forensic DNA typing (3.ª ed.). Elsevier.

Ballard, D., Winkler-Galicki, J., y Wesoły, J. (2020). Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. International Journal of Legal Medicine, 134(4),

1291–1303.

Bewes, J., Low, A., Morphett, A., Pate, F. D., y Henneberg, M. (2019). Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. Journal of Forensic and Legal Medicine, 62, 40–43.

Billah, M. (2025). Explainable AI for digital forensics: Ensuring transparency in legal evidence analysis. Journal of Forensic Science Research.

Brasil, Congresso Nacional. (2023). Projeto de Lei n.º 2338/2023 [marco legal da inteligência artificial].

Brożek, B., Furman, M., Jakubiec, M., y Kucharzyk, B. (2023). The black box problem revisited: Real and imaginary challenges for automated legal decision making. Artificial Intelligence and Law.

Chandra, N. A., Murtfeldt, R., Qiu, L., Karmakar, A., Lee, H., Tanumihardja, E., Farhat, K., Caffee, B., Paik, S., Lee, C., Choi, J., Kim, A., y Etzioni, O. (2025). Deepfake-Eval-2024: A multi-modal in-the-wild benchmark of deepfakes circulated in 2024 (arXiv:2503.02857). arXiv. https://arxiv.org/abs/2503.02857

Cho, S. H., Kim, D., Kwon, H. C., y Kim, M. (2024). Exploring the potential of large language models for author profiling tasks in digital text forensics. Forensic Science International: Digital Investigation.

Coble, M. D., y Bright, J. A. (2019). Probabilistic genotyping software: An overview. Forensic Science International: Genetics, 38, 219–224.

Cole, S. A. (2009). Forensics without uniqueness, conclusions without individualization: The new epistemology of forensic identification. Law, Probability and Risk, 8(3), 233–255. https://doi.org/10.1093/lpr/mgp016

Colliot, O. (2023). A non-technical introduction to machine learning. En O. Colliot (Ed.), Machine learning for brain disorders (pp. 1–31). Humana. https://doi.org/10.1007/978-1-0716-3195-9_1

Crawford, A. M., Ommen, D. M., y Carriquiry, A. L. (2023). A statistical approach to aid examiners in the forensic analysis of handwriting. Journal of Forensic Sciences, 68(5), 1531–1547.

Cunha, R. R., Arrabal, C. T., Dantas, M. M., y Bassanelli, H. R. (2022). Laser scanner and drone photogrammetry: A statistical comparison between 3-dimensional models and its impacts on outdoor crime scene registration. Forensic Science International, 330, 111100.

Departamento de Justicia de Estados Unidos, Oficina del Inspector General. (2006). A review of the FBI's handling of the Brandon Mayfield case. https://www.oversight.gov/sites/default/files/oig-reports/s1105.pdf

Elkins, K. M., Garloff, A. T., y Zeller, C. B. (2023). Additional predictions for forensic DNA

phenotyping of externally visible characteristics using the ForenSeq and Imagen kits. Journal of Forensic Sciences, 68(2), 608–613. https://doi.org/10.1111/1556-4029.15215

Epain, M., Valette, S., Zou, K., Ho Ba Tho, M. C., y Dumas, R. (2024). Sex estimation from coxal bones using deep learning in a population balanced by sex and age. International Journal of Legal Medicine, 138, 2617–2623. https://doi.org/10.1007/s00414-024-03268-2

Explainable AI and law: An evidential survey. (2023). Digital Society, 3(1). https://doi.org/10.1007/s44206-023-00081-z

Foley, M. M., y Oldoni, F. (2023). A global snapshot of current opinions of next-generation sequencing technologies usage in forensics. Forensic Science International: Genetics, 63, 102819.

Forensic Statistics Working Group. (2019). Influencing forensic science for a decade: Remembering the 2009 NAS report. https://forensicstats.org/blog/influencing-forensic-science-for-a-decade-remembering-the-2009-nas-report

Friedl, S., y Pernul, G. (2024). IoT forensics readiness: Influencing factors. Forensic Science International: Digital Investigation, 49, 301768. https://doi.org/10.1016/j.fsidi.2024.301768

García Morales, F. (2021). Daubert para la admisibilidad de la prueba técnica y científica en los procesos judiciales y su modulación. Diálogo Forense, 2(4), 34–40.

Garrett, B. L. (2021). Autopsy of a crime lab: Exposing the flaws in forensics. University of California Press.

Georgiou, A., Masters, P., Johnson, S., y Feetham, L. (2022). UAV-assisted real-time evidence detection in outdoor crime scene investigations. Journal of Forensic Sciences, 67(3), 1221–1232. https://doi.org/10.1111/1556-4029.15009

Girard, J. E. (2021). Criminalistics: Forensic science, crime, and terrorism (5.ª ed.). Jones & Bartlett Learning.

Greenwood, S. M., Paduch, C. D., y Allen, T. T. (2023). An evaluation of measurement uncertainty of trajectory angles using a 3D laser scanner. Journal of Forensic Sciences, 68(3), 828–839.

Grosz, S. A., y Jain, A. K. (2023). Latent fingerprint recognition: Fusion of local and global embeddings. IEEE Transactions on Information Forensics and Security, 18, 5691–5705.

Grother, P., Ngan, M., y Hanaoka, K. (2019). Face recognition vendor test (FRVT) part 3: Demographic effects (NISTIR 8280). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

Guerrini, C. J., Kalokairinou, L., Robinson, J. O., Bash Brooks, W., Fullerton, S. M., Huston, S., Dahlquist, J., Madden, D., Crossnohere, N., Campoamor, N., Bridges, J. F. P., y McGuire, A. L. (2025). Investigative genetic genealogy practices warranting policy attention: Results of a modified policy Delphi. PLOS Genetics, 21(1), e1011520. https://doi.org/10.1371/journal.pgen.1011520

Guyll, M., Madon, S., Yang, Y., Burd, K. A., y Wells, G. (2023). Validity of forensic cartridge-case comparisons. Proceedings of the National Academy of Sciences, 120(20), e2210428120. https://doi.org/10.1073/pnas.2210428120

Herke, C. (2026). Automated fingerprint identification: The role of artificial intelligence in crime scene investigation. Forensic Sciences, 6(1), 6. https://doi.org/10.3390/forensicsci6010006

Hunt, J. (2018). Scientific validity and error rates: A short response to the PCAST report. Fordham Law Review, 86(5), 24–36.

Innocence Project. (2019, 18 de febrero). Ten years later: The lasting impact of the 2009 NAS report. https://innocenceproject.org/news/lasting-impact-of-2009-nas-report/

International Organization for Standardization. (2018–2025). ISO 21043: Forensic sciences (Partes 1–5). ISO.

INTERPOL. (2025). Interpol review of digital evidence 2023–2025. Forensic Science International: Synergy.

Janiesch, C., Zschech, P., y Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., y Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

Kassin, S. M., Dror, I. E., y Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. Journal of Applied Research in Memory and Cognition, 2(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Ketsekioulafis, I., Filandrianos, G., Katsos, K., Thomas, K., Spiliopoulou, C., Stamou, G., y Sakelliadis, E. I. (2024). Artificial intelligence in forensic sciences: A systematic review of past and current applications and future perspectives. Cureus, 16(9), e70363. https://doi.org/10.7759/cureus.70363

Klasén, L., Fock, N., y Forchheimer, R. (2024). The invisible evidence: Digital forensics as key to solving crimes in the digital age. Forensic Science International, 362, Article 112133. https://doi.org/10.1016/j.forsciint.2024.112133

Kottner, S., Thali, M. J., y Gascho, D. (2023). Using the iPhone's LiDAR technology to capture 3D forensic data at crime and crash scenes. Forensic Imaging, 32, 200535. https://doi.org/10.1016/j.fri.2023.200535

Kukucka, J., y Dror, I. E. (2023). Human factors in forensic science: Psychological causes of bias and error. En D. DeMatteo y K. Scherr (Eds.), The Oxford handbook of psychology and law (pp. 621–636). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197649138.013.36

Lázaro Ruiz, E. (2024). La Academia Mexicana de Criminalística: Un referente histórico y

científico de los servicios periciales. Revista Penal México, (24), 109–130.

Le Bouthillier, M. E., Hrynkiw, L., Beauchamp, A., y Ratté, S. (2023). Automated detection of regions of interest in cartridge case images using deep learning. Journal of Forensic Sciences, 68(4), 1190–1201.

Lee, Y., Bradford, B., y Posch, K. (2024). The effectiveness of big data-driven predictive policing: Systematic review. CrimRxiv. https://www.crimrxiv.com/pub/nc4nlhqi

Li, H. (2023). Handwriting identification and verification using artificial intelligence-assisted textural features. Scientific Reports, 13, 21739. https://doi.org/10.1038/s41598-023-48789-9

Lidén, M., Thiblin, I., y Dror, I. E. (2023). The role of alternative hypotheses in reducing bias in forensic medical experts' decision making. Science & Justice, 63(5), 581–587. https://doi.org/10.1016/j.scijus.2023.07.005

Locard, E. (1920). L'enquête criminelle et les méthodes scientifiques. Ernest Flammarion.

Lum, K., y Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19.

Maltoni, D., Maio, D., Jain, A. K., y Feng, J. (2022). Handbook of fingerprint recognition (3.ª ed.). Springer Nature.

Melkote, M., y Nambiar, M. (2025). Access to algorithmic justice: Defending a right to expert assistance for indigent defendants. Duke Law Journal Online, 74, 63–79.

Mikros, G. (2025). Large language models and forensic linguistics: Navigating opportunities and threats in the age of generative AI (arXiv:2512.06922). arXiv. https://arxiv.org/abs/2512.06922

Mohd Sabri, N. E., Chainchel Singh, M. K., Mahmood, M. S., Khoo, L. S., Mohd Yusof, M. Y. P., Heo, C. C., Muhammad Nasir, M. D., y Nawawi, H. (2023). A scoping review on drone technology applications in forensic science. SN Applied Sciences, 5(9), 1–14. https://doi.org/10.1007/s42452-023-05450-4

Monson, K. L., Smith, E. D., y Peters, E. M. (2022). Accuracy of comparison decisions by forensic firearms examiners. Journal of Forensic Sciences, 68(1), 86–100. https://doi.org/10.1111/1556-4029.15152

Moreno, A. B., y Silva, C. D. (2020). Historia y evolución de la criminalística: Del método Bertillon a la era digital. Archivos de Criminología, 18(1), 12–28.

National Institute of Justice. (2022). Crime scene documentation: Weighing the merits of three-dimensional laser scanning. https://nij.ojp.gov/topics/articles/crime-scene-documentation-weighing-merits-three-dimensional-laser-scanning

National Research Council. (2009). Strengthening forensic science in the United States: A path forward. The National Academies Press. https://doi.org/10.17226/12589

Nini, A. (2023). A theory of linguistic individuality for authorship analysis. Cambridge

University Press.

Office of Justice Programs. (2019). Needs assessment of forensic laboratories and medical examiner/coroner offices: Report to Congress. U.S. Department of Justice.

Páez, A. (2021). Negligent algorithmic discrimination. Law and Contemporary Problems, 84(3), 19–33.

Parlamento Europeo y Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial). Diario Oficial de la Unión Europea.

Pei, G., Zhang, J., Hu, M., Zhang, Z., Wang, C., Wu, Y., Zhai, G., Yang, J., Shen, C., y Tao, D. (2024). Deepfake generation and detection: A benchmark and survey (arXiv:2403.17881). arXiv. https://arxiv.org/abs/2403.17881

Perú, Congreso de la República. (2023). Ley N.° 31814, Ley que promueve el uso de la inteligencia artificial (IA) en favor del desarrollo económico y social del país.

President's Council of Advisors on Science and Technology. (2016). Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. Executive Office of the President.

President's Council of Advisors on Science and Technology. (2017). An addendum to the PCAST report on forensic science in criminal courts. Executive Office of the President.

Qureshi, S. M., Saeed, A., Almotiri, S. H., Ahmad, F., y Al Ghamdi, M. A. (2024). Deepfake forensics: A survey of digital forensic methods for multimodal deepfake identification on social media. PeerJ Computer Science, 10, e2037. https://doi.org/10.7717/peerj-cs.2037

Rajarajeshwari, B. S., y Architha, M. R. (2025). AI transformation: Student learning assessment, skill development and professional readiness in science, forensics and cyber education: A review. Atlantis Press.

Russell, S., y Norvig, P. (2021). Artificial intelligence: A modern approach (4.ª ed.). Pearson.

Saferstein, R. (2020). Criminalistics: An introduction to forensic science (13.ª ed.). Pearson.

Sazaly, A. N., Mohd Ariff, M. F., y Razali, A. F. (2023). 3D indoor crime scene reconstruction from micro UAV photogrammetry technique. Engineering, Technology & Applied Science Research, 13(6), 12020–12025.

Scurich, N., Albright, T. D., Stout, P., Eudaley, D., Neuman, M., y Hundl, C. (2025). The Hawthorne effect in studies of firearm and toolmark examiners. Journal of Forensic Sciences. https://doi.org/10.1111/1556-4029.70047

Sributsayakarn, N., Intharah, T., Hirunchavarod, N., Pornprasertsuk-Damrongsri, S., y Jirarattanasopha, V. (2025). Automated age and sex estimation from dental panoramic radiographs.

International Dental Journal, 75(6). https://doi.org/10.1016/j.identj.2025.103967

State v. Puloka, No. 21-1-04851-2 (Wash. Super. Ct. Kings County 2024).

Sunde, N., y Dror, I. E. (2019). Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. Digital Investigation, 29, 101–108.

Sunde, N., y Dror, I. E. (2021). A hierarchy of expert performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. Forensic Science International: Digital Investigation, 37, 301175.

Swofford, H., Lund, S., Iyer, H., Butler, J., Soons, J., Thompson, R., Desiderio, V., Jones, J. P., y Ramotowski, R. (2024). Inconclusive decisions and error rates in forensic science. Forensic Science International: Synergy, 8, 100472.

Sánchez, J. S. V. (2021). Exclusión de la prueba pericial científica (de baja calidad epistémica) en fase de admisibilidad en procesos penales de tradición románica-continental: Diálogo entre dos culturas jurídicas. Revista Brasileira de Direito Processual Penal, 7(1), 375–408. https://doi.org/10.22197/rbdpp.v7i1.498

Taboga, J. (2022). Exploring the criminal procedure reform processes: Analytical perspectives on the reconfiguration of criminal justice in Latin America. Delito y Sociedad, (53). https://doi.org/10.14409/dys.2022.53.e0054

Terrado-Ortuño, N., y May, P. (2024). Forensic DNA phenotyping: A review on SNP panels, genotyping techniques, and prediction models. Forensic Sciences Research, 9(1). https://doi.org/10.1093/fsr/owae013

Verdú, F. (2023). Prueba pericial y prueba científica: Sus diferencias. Gaceta Internacional de Ciencia Forense, (49), 1–13.

Vestal, C. (2025, 21 de julio). Forensic crime labs are buckling as new technology increases demand. Stateline.

Volonnino, G., De Paola, L., Spadazzi, F., Serri, F., Ottaviani, M., y Zamponi, M. V. (2024). Artificial intelligence and future perspectives in forensic medicine: A systematic review. Clinica Terapeutica, 175(3), 193–202. https://doi.org/10.7417/CT.2024.5062

Wang, X., Liu, G., Wu, Q., Zheng, Y., Song, F., y Li, Y. (2024). Sex estimation techniques based on skulls in forensic anthropology: A scoping review. PLoS ONE, 19(12), e0311762. https://doi.org/10.1371/journal.pone.0311762

Whitehead, F. A., Williams, M. R., y Sigman, M. E. (2023). Analyst and machine learning opinions in fire debris analysis. Forensic Chemistry, 35, 100517. https://doi.org/10.1016/j.forc.2023.100517

Xu, Z., Jain, S., y Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models (arXiv:2401.11817). arXiv. https://arxiv.org/abs/2401.11817



MGS. JOSÉ LUIS ERAZO PARRA

Universidad Nacional de Chimborazo

jlerazo@unach.edu.ec

https://orcid.org/0000-0003-3149-6718

Mgs. José Luis Erazo Parra, profesional con una destacada trayectoria en la educación superior y la gestión tecnológica. Es Ingeniero en Sistemas e Informática y Magíster en Informática Empresarial por la Universidad Regional Autónoma de los Andes (UNIANDES). Con más de diez años de experiencia académica en instituciones como la Escuela Superior Politécnica de Chimborazo (ESPOCH) y la Universidad Nacional de Chimborazo (UNACH), ha dictado cátedras en informática, metodología de la investigación y TIC. A lo largo de su carrera, ha liderado procesos directivos y de coordinación académica, impulsando la innovación tecnológica como eje fundamental para el desarrollo y fortalecimiento institucional.



MGS. ELIANA DEL CONSUELO DE LA TORRE NÚÑEZ

Universidad Estatal de Bolivar
eliana.delatorre@ueb.edu.ec
https://orcid.org/0000-0001-8458-6632

Mgs. Eliana del Consuelo de la Torre Núñez, investigadora y docente en la Universidad Estatal de Bolívar (UEB). Posee una diversa formación que integra el Laboratorio Clínico e Histopatológico, la Ingeniería Ambiental y una Maestría en Criminalística y Ciencias Forenses. Su trayectoria contempla 20 años de labor académica e investigativa en la Universidad Nacional de Chimborazo en las áreas de laboratorio y gestión. Ha participado en múltiples proyectos multidisciplinarios y publicaciones científicas, posicionándose como un referente comprometido con la innovación pedagógica y el rigor metodológico en las ciencias de la salud y forenses.



MGS. FREDDY RODOLFO SÁNCHEZ PÓLIT

Universidad Estatal de Bolivar
freddy.sanchez@ueb.edu.ec
https://orcid.org/0009-0002-2569-9906

Mgs. Freddy Rodolfo Sánchez Polit es un profesional con una destacada trayectoria en las ciencias forenses, el peritaje judicial y la docencia universitaria. Es Magíster en Criminalística y Ciencias Forenses y se desempeña como oficial de la Policía Nacional del Ecuador, acumulando más de doce años de experiencia en la investigación técnico-científica. Especializado en la escena del crimen, análisis digital y balística, ha liderado procedimientos complejos bajo rigurosos estándares internacionales y de preservación probatoria. Actualmente, ejerce la docencia en la Universidad Estatal de Bolívar (UEB), aportando al ámbito académico una sólida visión práctica, operativa y de rigor en la criminalística.



MGS. VERÓNICA PAULINA CÁCERES MANZANO

Universidad Nacional de Chimborazo

vcaceres@unach.edu.ec

https://orcid.org/0000-0001-5710-5661

MsC. Verónica Paulina Cáceres Manzano es una profesional multidisciplinaria con formación en Ciencias de la Salud (Laboratorio Clínico e Histopatológico, UNACH), Derecho (Universidad Regional Autónoma de los Andes), y posgrados en Criminalística (Universidad Internacional del Ecuador) y Gestión de la Seguridad Clínica (Universidad Internacional de La Rioja, España). Con una década de experiencia como perito forense en el Servicio Nacional de Medicina Legal y Ciencias Forenses, y como docente en la Universidad Nacional de Chimborazo, ha publicado numerosos artículos científicos en áreas como biología forense, odontología forense y análisis clínicos. Su trabajo combina investigación, docencia y aplicación práctica en el ámbito forense.



DRA. MARÍA EUGENIA LUCENA DE USTÁRIZ

Universidad Nacional de Chimborazo

mlucena@unach.edu.ec

https://orcid.org/0000-0001-9120-345X

Dra. María Eugenia Lucena de Ustáriz es una destacada investigadora y docente con formación en Bioanálisis (Licenciatura, Universidad de Los Andes), Maestría en Biotecnología y PhD en Ingeniería de Procesos (Universidad de Oviedo). Con más de 25 años de experiencia académica en Venezuela y Ecuador, actualmente ejerce como docente-investigador en la Universidad Nacional de Chimborazo. Su investigación se centra en productos naturales, bioactividad y análisis forense, respaldada por numerosas publicaciones científicas y proyectos en colaboración. Su labor integra docencia, investigación y vinculación comunitaria, destacando su contribución al avance científico y educativo.



DR. FRANCISCO JAVIER USTÁRIZ FAJARDO

Universidad Nacional de Chimborazo
francisco.ustariz@unach.edu.ec
https://orcid.org/0000-0002-6423-9067

Dr. Francisco Javier Ustáriz Fajardo es un destacado académico e investigador con formación en Bioanálisis (Licenciatura, Universidad de Los Andes, Venezuela), Maestría en Biotecnología de Microorganismos y Doctorado en Tecnología del Medio Ambiente (Universidad de Oviedo, España). Con más de 25 años de experiencia docente e investigadora en universidades de Venezuela y Ecuador, ha publicado numerosos artículos científicos en áreas como microbiología, biotecnología y farmacología. Actualmente, se desempeña como docente-investigador en la Universidad Nacional de Chimborazo, Ecuador, contribuyendo a la formación en ciencias de la salud.

La criminalística atraviesa una de las transformaciones más profundas de su historia. La inteligencia artificial, la biometría avanzada, la genética forense de nueva generación, la reconstrucción tridimensional de escenas, el análisis algorítmico de rastros y la evidencia digital están redefiniendo la manera en que se identifican, interpretan y presentan los indicios ante la justicia.

Criminalística del futuro: Ciencia, inteligencia artificial y transformación forense examina este nuevo escenario desde una perspectiva científica, jurídica y ética. A lo largo de diez capítulos, la obra recorre la evolución de la disciplina desde los principios de Locard hasta los actuales sistemas inteligentes, analizando sus aplicaciones en identificación humana, genética forense, reconocimiento facial, balística, documentoscopia, informática forense, deepfakes, criptoactivos, cadena de custodia digital y prueba algorítmica.

El libro presta especial atención a los desafíos de validez científica, sesgo cognitivo y algorítmico, explicabilidad, privacidad, derechos humanos, admisibilidad probatoria y estándares de calidad, evitando presentar la tecnología como una solución automática o infalible.
Con un enfoque especialmente dirigido al contexto iberoamericano, esta obra está pensada para peritos, criminalistas, fiscales, jueces, defensores, investigadores, docentes y estudiantes que necesitan comprender cómo utilizar las nuevas tecnologías sin abandonar el rigor científico que debe sostener toda prueba forense.

ISBN: 978-9907-9622-6-0