REVISTA CHILENA
DE DERECHO Y TECNOLOGÍA

VOL. 13 (2024) • e72923 • DOI 10.5354/0719-2584.2024.72923
RECIBIDO: 30/11/ 2023 • APROBADO: 8/12/2024 • PUBLICADO: 31/12/2024

DOCTRINA

Herramientas predictivas del riesgo de reincidencia criminal basadas en inteligencia artificial: Hacia su compatibilidad con la intimidad y la defensa del procesado en el juicio oral

Predictive tools for the risk of criminal recidivism based on artificial intelligence: Towards their compatibility with the privacy and defense of the accused in the oral trial

Manuel Urzúa Urzúa
Universidad de Talca, Chile

RESUMEN La inteligencia artificial aplicada al proceso judicial ha tenido como uno de sus elementos centrales la introducción de *softwares* de herramientas predictivas del riesgo de reincidencia criminal en diversas etapas del *iter* procesal. Desde su incorporación, se han ido dibujando los principales conflictos que riñen con una diversidad de derechos fundamentales, como el derecho fundamental a la intimidad y el derecho fundamental de defensa del procesado. En este trabajo, enfocado en la etapa de juicio oral, se ofrece una explicación sobre los principales nudos de esta compleja relación; se dibuja un esquema de configuración de su vulneración, abordándose el impacto procesal y, con base en lo anterior, se proponen posibles condiciones para compatibilizar su incorporación, sin que ello se traduzca en una vulneración de los mentados derechos.

PALABRAS CLAVE Inteligencia artificial, herramientas predictivas, intimidad, defensa.

ABSTRACT Artificial intelligence applied to the judicial process, has had as one of its central elements the introduction of software predictive tools for the risk of criminal recidivism at various stages of the procedural iter. Since its incorporation, the main conflicts that dispute with a variety of fundamental rights, such as the fundamental right to privacy and the fundamental right of defense of the indicted have been shown. This paper, focus in the oral trail stage, provides an explanation of the main nodes of this complex relationship; A configuration diagram of its violation is portrayed, addressing the procedural impact, and based on the above, possible conditions are proposed to make its incorporation compatible, without resulting in a violation of the aforementioned rights.

KEYWORDS Artificial Intelligence, predictive tools, privacy, defense.

Introducción

La inteligencia artificial (IA) ha superado las épocas de invierno y vive resueltamente una de pleno verano. A raíz de ello, existe una expectación internacional por la inminente regulación normativa de esta tecnología, esperándose una solución que concilie la protección, el desarrollo y la innovación. Es necesario abordar óptimamente los impactos e intromisiones en la esfera de diversos derechos fundamentales que se encuentran en una tensa relación con la IA, debido a la forma en cómo esta se produce y funciona; y, a su vez, canalizar sus innegables aportes en las más diversas disciplinas y campos del conocimiento humano. En dicho escenario, urge hallar un equilibro que permita transitar desde el espacio de los problemas hacia el de las soluciones, ante el riesgo de dilatar inexplicablemente su incorporación al proceso judicial. En el horizonte aparecen, a nivel comparado, la recientemente aprobada Artificial intelligence act del Parlamento Europeo,[1] o las directrices adoptadas por Estados Unidos,[2] instruidas por el presidente Joe Biden. A su turno, en Chile, con fecha 07 de mayo de 2024, se ingresó un proyecto de ley[3] que busca regular los sistemas de IA, adoptando enfoques de riesgo, tal como lo han hecho los instrumentos indicados previamente.

Para la doctrina procesal, se vislumbra una relación entre la IA y los derechos fundamentales que no es sencilla. El proceso criminal es un asunto de suyo delicado, pues se comprometen derechos fundamentales como la presunción de inocencia o la intimidad, de modo que esta tardanza en la regulación e incorporación no es un fenómeno *per se* negativo, sino que, por el contrario, ha dado lugar a una reflexión profunda con sentido ético antes de su total inmersión en el proceso. Entre los factores causantes de esta reflexión, uno importante es la experiencia reciente de uso de IA en el proceso criminal estadounidense. Esta ha permitido identificar diversos riesgos que conlleva la utilización de IA y, además, ha potenciado a nivel internacional el debate sobre posibles soluciones. Este trabajo solo aborda uno de los tantos problemas asociados a la posibilidad de utilizar herramientas predictivas del riesgo de reincidencia criminal en la fase de juicio oral, y su compleja compatibilidad con el derecho fundamental a la intimidad desde la óptica del derecho fundamental de defensa del procesado.[4]

1. Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, del 13 de mayo de 2024.

2. Executive order on the Safe, Secure, and Trustworsthy Development and Use of Artificial Intelligence, del 30 de octubre de 2023. Disponible en https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

3. Véase Boletín 16821-19, en primer trámite constitucional. Disponible en https://tipg.link/R-Sy.

4. En este trabajo, el término procesado dice relación con la persona que está siendo sometida a un proceso penal, específicamente, en la etapa de juicio oral, comprendiendo indistintamente términos como reo, acusado, imputado o investigado, dado que no desvían el objetivo propuesto por el autor.

A pesar de producirse con predominancia en Estados Unidos, la experiencia práctica de utilización de herramientas predictivas ha abierto el debate sobre su utilización en los diversos países del orbe. En dicho contexto, a pesar de la amplitud de los sistemas normativos existentes y su regulación interna, considero pertinente invitar al lector a efectuar una mirada global del problema planteado, pues, tal como reconoce el proyecto de ley que pretende regular la IA en nuestro país, o las consideraciones que anteceden la reciente ley de IA de la Unión Europea (UE), es un consenso[5] que la IA es un fenómeno con alcance internacional, y que su regulación —se insta a ello— debe ser adoptada con un enfoque basado en riesgos, siendo sus virtudes y potenciales afecciones homólogos en los distintos países. Una muestra de lo anterior es el hecho de que, por ejemplo, la definición de lo que se considera IA[6] es igual en ambos documentos. Sobre lo anterior, se previene que, en las partes que consideré pertinentes, aclaré ciertos aspectos normativos a la luz de la regulación europea, bajo utilidad de haberse seguido en nuestro país, tanto para la Política Nacional de IA como para el reciente proyecto de ley, aquellas directrices que ha seguido la UE, hoy materializadas en dicha normativa. Sin embargo, vale indicar que, matices más o matices menos, el proceso judicial es uno.

IA en el proceso criminal

A pesar de que definir la IA es una cuestión compleja (Amunátegui, 2020: 13), se entiende por ella toda referencia a «máquinas o agentes capaces de observar su entorno, de aprender, y basados en el conocimiento y la experiencia adquirida, de tomar acciones inteligentes o proponer decisiones» (UE, 2018: 18). Para adentrarnos en su estudio, es necesario delimitar con claridad el campo a investigar y, sumado a ello, el momento de desarrollo que atraviesa su aplicación en el proceso criminal, pues, por más reflexiones y expectativas sobre el uso de la IA en el derecho, distintos factores, como la no recopilación de datos (Esparza, 2022: 183 y 187); la relación delicada con distintos derechos fundamentales, como la defensa (Bueno de Mata, 2020: 27; Ward, 2020: 1); la intimidad (Nieva, 2018: 142) y la presunción de inocencia (Faggiani, 2022: 523); la ausencia de regulación normativa plenamente vigente o el estancamiento tecnológico (Nieva,

5. El proyecto dispone: «Así, de forma similar a la Ley de IA europea y a partir del consenso existente a nivel comparado en torno a la conveniencia de adoptar un enfoque basado en riesgos para la regulación de las tecnologías basadas en sistemas de IA, este proyecto de ley adopta dicho enfoque vinculado al desarrollo, implementación y uso de sistemas de IA».

6. La Ley de IA (UE) y el proyecto presentado al Congreso Nacional contienen la misma definición: «La inteligencia artificial (IA) es un conjunto de sistemas basado en máquinas que infieren, a partir de información de entrada, determinada información de salida, que puede consistir en predicciones, contenidos, recomendaciones o decisiones capaces de influenciar espacios físicos o virtuales». Respectivamente, artículo 3 número 1 de la normativa europea y párrafo número 1 de los «Antecedentes» del proyecto.

2022b: 417) han provocado que su incorporación sea solo una proyección más bien teórica que real.

Así, como punto de partida y evidencia de conflicto consumado podemos analizar la experiencia de aplicación de IA en Estados Unidos, dado que algunos Estados (Kehl y otros, 2017: 10-11) contemplan la utilización de las herramientas predictivas del riesgo de reincidencia criminal en el proceso penal. Siguiendo a Miró (2018: 107), dichas herramientas dicen relación con sistemas que son potencialmente aplicables a procesos de toma de decisiones relacionados con la valoración del riesgo de reincidencia de reos involucrados en un proceso judicial. Para materializar lo anterior, es indispensable la consideración de los factores de riesgo que contribuyen a la evaluación, pues ellos determinan qué datos se requieren del individuo para efectuar la predicción. Así, con base en lo planteado por Quattrocolo (2022: 147 y 148), estos factores se clasifican en estáticos y dinámicos, variando su medición en atención a las condiciones del contexto, siendo crucial su ponderación en el resultado arrojado.

Si bien estas herramientas se utilizan en diversas instancias del proceso penal (*pre-trial*, libertad provisional o libertad bajo fianza), el estudio se centrará en la etapa de juicio oral, a raíz del escenario de especial vulnerabilidad en que se haya el procesado previo a dicha etapa, desde la óptica de su derecho fundamental a la intimidad, la profundización de ese estado ante las medidas intrusivas (y progresivas) de investigación que puede emplear el ente persecutor, y las posibles implicancias del uso de IA para el despliegue de su derecho de defensa. El riesgo de ello se explica porque, pretendiendo aportar una predicción, es posible que se desvíe por completo el objeto de esta etapa procesal debido a que, en circunstancias que el foco de atención debe estar en la práctica probatoria que maximiza el conocimiento de los hechos en favor del juzgador y los litigantes, acercándoles a la verdad, la predicción arrojada por la herramienta podría disfrazarse sutilmente de prueba, logrando superponerse a ella, a pesar de ir claramente por carriles separados, dado que una predicción no es homologable a un medio de prueba o, al menos, su naturaleza jurídica es discutible.

Con base en lo anterior, es necesario referirme al funcionamiento de las herramientas predictivas, las técnicas de IA que utilizan y, en definitiva, a por qué algunas son más problemáticas que otras. Respecto del primer elemento, son varias las generaciones de herramientas evaluativas que han existido en Estados Unidos. Distintos autores (Turner y otros, 2013: 1 y 2; Kehl y otros, 2017: 9) identifican al menos cuatro generaciones. La primera de ellas (desarrollada en la primera mitad del siglo XX), centrada en un juicio clínico no estructurado *(professional judgment alone),* implicaba un análisis informal de datos, transformado enseguida en un juicio profesional disponible para adoptar medidas sobre el sujeto analizado, siendo realizado por personal clínico o penitenciario (Andrews y Bonta, 2007: 1; Turner y otros, 2013: 1; Horcajo y otros, 2019: 42). Luego, a

comienzos de la década de 1970, los instrumentos de segunda generación *(evidence-based tools)* incorporaron el análisis de criterios estructurados, objetivos y basados en evidencia, considerándose factores estáticos de medición, como el historial delictivo del sujeto, su edad o sexo. Ello posibilitó un efecto discriminatorio, pues, al ser elementos no controlables por el delincuente, y descartándose los factores dinámicos como la buena conducta, o la ausencia de abuso de alcohol después del delito, se redujo la posibilidad de contrarrestar el nivel de riesgo arrojado por factores fijos (Andrews y Bonta, 2007: 3 y 4; Turner y otros, 2013: 1; Horcajo y otros, 2019: 42).

La tercera generación *(evidence-based and dynamic),* de fines de 1970 y principios de 1980, incorporó abiertamente factores de riesgo dinámicos de manera predominante, pues los niveles de predicción de la herramienta podían aumentar (Turner y otros, 2013: 2). Dichos factores se relacionaron directamente con las necesidades criminógenas del sujeto. El modelo de riesgo-necesidad-respuesta (RNR) es un gran ejemplo de estas herramientas, que identifican, según Turner (2013, 2), una fuerte relación entre la evaluación competente del riesgo de reincidencia, la identificación de las necesidades del delincuente y la efectividad del tratamiento. La cuarta generación *(systematic and comprehensive)*, por último, presenta como característica principal un enfoque integral y sistémico, pues, basándose en la tercera generación de herramientas, busca medir la reincidencia en base a factores de riesgo y características específicas del delincuente, incorporando factores personales que antes no eran medidos por las generaciones anteriores de herramientas.

En el mismo sentido, según cómo lo ha planteado Kehl y otros (2017: 9), esta cuarta generación se ha potenciado por la utilización de tecnologías que permiten el cruce de datos de una manera mucho más óptima que las herramientas de generaciones anteriores,[7] siendo la técnica más avanzada para ello la conocida como *machine learning,* o al menos, la que ha otorgado resultados más precisos (Lehr y Ohm, 2017: 710). Siguiendo a Murphy (2012: 2), dicha técnica consiste en un «conjunto de métodos que sirven para detectar de manera automática patrones en los datos, y luego usar los patrones descubiertos para predecir datos futuros o para realizar otros tipos de toma de decisiones, bajo incertidumbre (como lo sería planificar una próxima recogida de datos)». O bien, en los términos que plantea la *executive order* de octubre pasado, dice relación con un «conjunto de técnicas que se pueden utilizar para entrenar algoritmos de IA para mejorar el rendimiento de una tarea basada en datos».[8]

7. Algunos autores, incluso, plantean que esta sería la quinta generación de herramientas. En este sentido, Taxman y Dezember (2016: 11) y Garreth y Monahan (2020: 451).

8. Executive order on the Safe, Secure, and Trustworsthy Development and Use of Artificial Intelligence, del 30 de octubre de 2023. Sección número 3, letra t).

Solo a modo de ejemplo, dentro de las herramientas predictivas que existen, está la famosa COMPAS,[9] y otras como LRI- I o PSA.[10] Algunas utilizan técnicas simples de IA, presentando mayores virtudes de explicabilidad o interpretabilidad en sus resultados (Burrell, 2016: 9), y permitiendo eventualmente a los operadores jurídicos una mayor comprensión de ellos y del funcionamiento algorítmico que los produce. Sin embargo, parte de la doctrina es escéptica en este punto (Edwards y Veale (2017: 23) apuntan al riesgo de una «transparencia sin sentido» del algoritmo), pues la complejidad de la técnica colabora simultáneamente en una mayor precisión en la predicción. Así, la técnica de *machine learning* toma ventaja, pues su capacidad de analizar datos e identificar patrones es muy alta, siendo claramente una técnica satisfactoria en la tarea de lograr mejores predicciones, dado que se adapta dinámicamente a nuevos datos (Kehl y otros, 2017: 9), cuestión necesaria para seguir entrenando a la herramienta respectiva y, además, para seguir manteniendo su capacidad predictiva frente a nuevos sujetos procesados y a los factores dinámicos propios de estas personas. En otros términos, para seguir desarrollando una «predicción en tiempo real» (Garreth y Monahan, 2020: 451).

Como resultado de lo anterior, mi trabajo sitúa en un espacio común aquellas herramientas predictivas que son utilizadas en la fase de juicio oral; que presentan algoritmos de carácter privado; y, en último término, que utilizan predominantemente la técnica de *machine learning* para ser diseñadas, entrenadas y finalmente utilizadas.

En búsqueda de la transparencia

El primer conflicto transcendental de la IA en el proceso criminal dice relación con la ausencia de transparencia del funcionamiento de las herramientas predictivas. Aquí, es indispensable apuntar que el elemento central de toda IA son sus algoritmos. Según Gillespie (2014: 167), los algoritmos son «procedimientos codificados para transformar datos de entrada en una salida deseada, en función de cálculos especificados», constituyendo la clave de la IA. Para Quattrocolo (2022: 9), la utilización de estos se vincula a fenómenos que podrían conducir a decisiones no discrecionales.

9. Correctional Offender Management Profiling for Alternative Sanctions (administración de perfiles de criminales para sanciones alternativas del sistema de prisiones) es una herramienta creada por la empresa norteamericana Northpointe (hoy rebautizada como Equivant) y fue la utilizada en el caso *State of Wisconsin vs. Loomis*, 881 N.W.2d 749 (Wis. 2016), de repercusión global.

10. LSI-R es una de las primeras herramientas predictivas, desarrollada por Multi-Health Systems. Extrae información de una encuesta que contiene factores estáticos y dinámicos, que van desde antecedentes penales hasta patrones de personalidad. Se utiliza en fase de sentencia, no obstante, la herramienta fue desarrollada inicialmente para su uso en rehabilitación. Hoy se utiliza en Washington y California. En detalle, véase Kehl y otros, 2017: 11. Por su parte, Public Safety Assessment se utiliza en el *pre-trial,* en al menos 28 jurisdicciones de Estados Unidos. En detalle, véase Quattrocolo, 2022: 151 y 152.

Si existe alguna característica que se ha imputado a la IA, y en particular a la técnica de machine learning, es sin duda la black box que hay detrás de sus procedimientos o resultados. En dicho sentido, se ha hablado igualmente de la «inescrutabilidad de la técnica» (Lehr y Ohm, 2017: 656), pues la técnica de machine learning no permite saber o averiguar el procedimiento seguido por ella para llegar a un resultado, señal de su complejidad técnica.

Sobre esto, hay quienes previenen en un doble sentido. Nieva (2022b: 419) plantea que es bueno cuidarse del «lenguaje algo falaz» a propósito de un abundante grupo de terminologías propias del lenguaje informático, como el *deep learning* o las redes neuronales, pues detrás de ellas se escondería que lo único que hace la IA es agrupar una cantidad de datos considerable, contrastándolos con los que ya posee. Otros, como Lehr y Ohm (2017: 669), plantean que la delimitación clara de estos términos podría colaborar con la búsqueda de soluciones en materia jurídica. Cualquiera sea el caso, el derecho procesal se halla en el imperativo de familiarizarse con un lenguaje que no había sido necesario utilizar antes.[11] Propio del campo de la ingeniería computacional y la informática, el lenguaje algorítmico permitirá progresivamente un mejor diálogo con la IA y sus desarrolladores.[12] Y es que, una lección a ponerse en práctica es aquella que llama a generar un espacio transversal e interdisciplinario de estudio e investigación sobre la materia (Hildebrant, 2018: 7; Nieva, 2022b: 418), pues, de lo contrario, no habrá soluciones integrales.

En esa línea, si de buscar la transparencia se trata, es inevitable hablar de la opacidad, entendiendo por ella una «incomprensibilidad remediable» (Pasquale, 2015: 15) del funcionamiento de los algoritmos. Con ella, se refiere la falta de transparencia motivada de la *black box*, pues está ausente una capacidad explicativa de la máquina, y se dificulta la inteligibilidad respecto del procedimiento seguido para llegar a un resultado determinado, una vez que se procesan los datos necesarios para que la máquina funcione.

Lo anterior conduce a que, si bien es relevante el tratamiento específicamente procesal de las tres etapas de funcionamiento de una herramienta basada en *machine learning* —que serán siempre tres etapas: incorporación de datos a la máquina (datos de entrada o *inputs*); procesamiento de los mismos (el algoritmo entra en acción, aplicando lo aprendido previamente en el proceso de diseño y entrenamiento); y en última instancia, el pronóstico de reincidencia (datos de salida u *outputs*)—, y virtualmente constituye el aspecto más atractivo de analizar, no debe abandonarse la mirada a las etapas previas que debe sortear una herramienta antes de aterrizar finalmente

11. De hecho, la novedad que ha traído la IA no es que se usen herramientas predictivas, sino que dicha tarea sea entregada a un *software* de IA. Véase Kehl y otros (2017: 2) y Dressel y Farid (2018: 1).

12. Muestra de ello, es que la jueza Shirley Abrahamson de la Corte Suprema de Wisconsin reconoció que la falta de compresión sobre el funcionamiento de COMPAS fue un problema significativo para la Corte, en *State vs. Loomis*, 881 N.W.2d 774 (2016), p. 133.

en el proceso criminal. En otros términos, saber quién, cómo y cuándo se produce la herramienta, es vital, pues se trata de variadas etapas «pre-procesales» que responden al desarrollo tecnológico de la misma, y que incluso pueden tener una mayor trascendencia, en cuanto a opacidad se refiere, que el período posterior de funcionamiento judicial, donde solo se manifiesta o revela este aspecto como una cuestión problemática. En suma, la opacidad no nace en el proceso, sino que solo cobra vida en él, obligándonos, a las y los juristas, a ir más allá para descubrir su raíz.

Bajo el alero de lo anterior, la opacidad presenta distintos niveles y espacios de manifestación. Un primer nivel, de carácter legal,[13] está expresado en la propiedad intelectual del desarrollador de la IA, cuya consecuencia es la inaccesibilidad al algoritmo, prohibición con efecto *erga omnes* que desliza una casi imposible conciliación entre el derecho de defensa del procesado y el secreto empresarial[14] debido a que resultará una especie de condición irrenunciable su no revelación, so pena de perder ventajas competitivas en el mercado (Burrell, 2016: 3 y Kehl y otros, 2017: 28). En este sentido, arrastrar la protección al secreto desde la regulación del derecho privado al derecho procesal, parece inapropiado. Incluso, Wexler (2018: 1395 y 1396) va más allá, y plantea que una protección del secreto empresarial, en esas condiciones, implica una sobreprotección de la propiedad intelectual, que socava la legitimidad del proceso criminal. La relación entre la propiedad intelectual y aspectos como la creación de perfiles, la opacidad, los errores, los sesgos y la eventual falta de justicia en el uso de aplicaciones de *machine learning*, según Azuaje (2023: 17), está explicada en un doble aspecto. En primer lugar, dichas aplicaciones normalmente acuden al régimen jurídico de la propiedad intelectual para resguardar sus derechos y, en segundo lugar, porque diversos instrumentos internacionales, como la ley de IA de la UE (artículo 86), apelan a una IA que tenga como principio irrenunciable la explicabilidad de las decisiones que adopte o sugiera, siendo muestra ello de que se configura como un conflicto específico en materia de IA y derechos fundamentales.

Un segundo nivel, de carácter técnico, está compuesto esencialmente por aquella complejidad que existe en el funcionamiento algorítmico. En este sentido, no basta respecto de la transparencia que el productor de la herramienta revele el algoritmo utilizado, pues la opacidad técnica sería inherente al funcionamiento de la *machine learning* (Edwards y Veale, 2017: 23). En el mismo sentido, indica Breiman (2001: 209), que sería imposible la comprensión de una «caja negra» que tiene cincuenta «árboles»

13. Lo que para Han-wei y otros (2018: 16 y 17) es una «legal black box», para Burrell (2016: 4 y 5) es un «primer nivel de opacidad».

14. He ahí la justificada preocupación de Chiao (2019: 137), en orden a que no se han podido visualizar garantías respecto de que las empresas desarrolladoras de IA en materia criminal respondan al interés público que reviste tanto la investigación criminal como el ejercicio jurisdiccional, lo que entronca con la visión de Carabantes (2023: 427-430) y el posible «secreto intencional» como segunda capa de opacidad.

enganchados entre sí, aludiendo con ello a los «árboles de decisiones», que, junto a las «redes neuronales», son modelos de algoritmos que presentan complejidades muy diferentes. El nacimiento de una rama específica de la IA, que pone su foco en la explicabilidad o interpretabilidad, responde a esa complejidad, según indica Solar (2022: 151), pauta seguida por la Ley de IA de la UE, e incluso por el proyecto de ley que pretende regular sistemas de IA en Chile.[15] Frente a este segundo nivel de opacidad, no hay demasiadas opciones, aparte de intentar salir del «analfabetismo técnico» (Burrell, 2016: 4).[16]

En el terreno pre-procesal, la culminación de un proceso de entrenamiento de una *machine learning* se materializará en la obtención del derecho de propiedad industrial sobre el algoritmo, profundizándose con ello, bajo mi perspectiva, otras formas de opacidad. Como lúcidamente se ha planteado (Kehl y otros, 2017: 28), la opacidad impide investigación y auditorías sobre estas herramientas.[17] Así, el ocultamiento no es un buen punto de partida. Y es que, todo comienza, desde la óptica procesal, con el candado de la propiedad industrial, un «privilege procesal» (Solar, 2022: 141) que, bajo el prisma de la normativa europea generaría una «grave cuestión de equilibro entre intereses legales contrapuestos» (Quattrocolo, 2002: 167), al contrario de lo sucedido en Estados Unidos, que ha validado su uso en esas condiciones. Es, quizás sea acertado decirlo, el problema primitivo de la opacidad.

En este escenario de doble opacidad, lo mínimo que se ha reclamado es la publicidad del algoritmo, pero en ningún caso clausurar anticipadamente el debate. La experiencia recogida a propósito de la utilización de COMPAS[18] en Estados Unidos no es más que un acontecimiento que permitió dimensionar los alcances de las consecuencias que una prohibición de acceso al algoritmo produce en los derechos fundamentales del procesado, pero también para evaluar el comportamiento del órgano jurisdiccional. Si bien no se ha paralizado su uso, con el riesgo «real y considerable» que su validación judicial conlleva (Nieva, 2022a: 92), se ha dejado el campo abierto para la investigación sobre la materia, persistiendo la contravención a la publicidad que significa el indicado secreto algorítmico (Nieva, 2018: 140; Martínez, 2020: 499) como un problema

15. En el caso de la UE, considerando 27 del acto y artículo 86, en el «derecho a explicación». En el caso del proyecto en Chile, es un principio expreso, recogido en el artículo 4, letra d).

16. Para Carabantes (2023: 427-430), la tercera capa de opacidad.

17. He ahí que, según plantea Azuaje (2023: 18), lo que se demande progresivamente respecto de ellas sea justamente la «explicabilidad, la transparencia o la auditabilidad».

18. La importancia del caso *State of Wisconsin vs. Loomis* (2016) radica en que la Corte Suprema de Wisconsin se pronunció extensamente sobre las «*risk assesment tools*», constituyendo un precedente jurisprudencial relevante para acercarse al funcionamiento de estos instrumentos predictivos e identificar cómo el órgano jurisdiccional interactúa con ellos y sus resultados. Sin embargo, para mayor claridad, COMPAS representa un software que utiliza algoritmos simples, no siendo homologable a la IA que se produce mediante *machine learning,* y que es a la que se presta más atención en el presente trabajo.

fundamental que condiciona todo.

En conclusión, los niveles, normativo y técnico se relacionan de modo tal que el primer nivel encubre al segundo; y en materia de espacios de manifestación, encontramos un espacio pre-procesal y un espacio puramente procesal, donde se produce y donde funciona la herramienta, respectivamente. La opacidad normativa contraviene garantías de publicidad y transparencia del proceso, constituyendo frente a la defensa un punto de tensión de origen pre-procesal. La opacidad técnica se despliega de manera distinta. Una primera etapa, propia de cómo y cuándo se produce la herramienta, contexto de problemas al parecer solo de índole técnica, con un impacto procesal suspendido, pues no se ha incorporado la herramienta al proceso; y una segunda, que devela todos los conflictos derivados de su secreto, pues no permite a los litigantes comprender, y con ello legitimar el resultado predicho, debido al desconocimiento sobre cómo se llega al resultado. Precisamente, en este concreto punto, la opacidad técnica se transforma en una opacidad procesal.

Los sesgos frente a la imparcialidad del juzgador

El segundo conflicto a tratar, que adquiere una mayor envergadura sobre la base de la opacidad, tiene como elemento nuclear el antecedente de que las herramientas predictivas pueden presentar sesgos al arrojar su predicción, pudiendo estos ser de diversa índole, siendo uno de los más graves el sesgo discriminatorio por razón de raza, sin ser excluyentes otros factores, como el género, las dificultades financieras o el recibir asistencia social (Hannah-Moffat y Avila, 2023: 552), que denotan la posibilidad de que se produzca una discriminación algorítmica.[19] En el contexto del proceso criminal, la utilización de herramientas predictivas con sesgos discriminatorios atenta, en el fondo, contra la imparcialidad del juzgador, siempre que la predicción de riesgo termine por influir de manera importante el fallo del tribunal, punto sobre el cual me explayaré más adelante. Sin embargo, desde ya es necesario advertir que los sesgos no son ajenos al ejercicio jurisdiccional,[20] de ahí que sería incorrecto exigir de una herramienta cuestiones que no han sido exigibles tampoco a los jueces humanos. Siguiendo a Quattrocolo (2022: 170) y a Hannah-Moffat y Avila (2023: 549), no todos los problemas que se han puesto en el tapete por el caso COMPAS derivan de la naturaleza digital de las herramientas predictivas del riesgo de reincidencia criminal, sino que son de larga data.

19. Los tipos de discriminación se han materializado en diversas situaciones, atendiendo a diferentes factores. Por ejemplo, es famoso el caso de la selección de currículos de Amazon mediante un agente de selección de personal a través de IA, siendo en dicho caso el género el factor de discriminación. Véase «Amazon scraps secret AI recruiting tool that showed bias against women», Reuters, octubre de 2018.

20. Tal cual como se apuntó respecto de las herramientas (nota 11), no hay novedad en el hecho de que existan sesgos, sino en su automatización.

Reconociendo en los sesgos un problema, la doctrina ha observado que puede conllevar la aparición de un posible doble efecto: uno indeseable y otro más auspicioso. El primero, el riesgo de «reproducir formas de opresión algorítmicas» (Hannah-Moffat y Avila, 2023: 549). Y un segundo efecto, que permitiría una identificación temprana de sesgos y que debiese conducir a una acción correctiva del funcionamiento de las herramientas —contrarrestando el primer efecto—, o incluso, yendo más allá, tal cual como se ha planteado con el «principio de no discriminación» en documentos de la UE,[21] asignar a la IA un rol para combatir activamente la discriminación,[22] identificándose un papel protagonista, y no para simplemente «evitar» situaciones de discriminación a raíz del funcionamiento algorítmico.

Las herramientas predictivas manifestarán sus sesgos al arrojarse la predicción de riesgo. Por ejemplo, en el sesgo por motivos de raza, la predicción podría otorgar una mayor cantidad de falsos positivos a personas afrodescendientes, en comparación a personas blancas.[23] Aquella denominación técnica hace referencia a situaciones en las que, habiéndose predicho por la herramienta un riesgo alto de reincidencia, esta predicción, con posterioridad, es calificable como errónea, pues finalmente las personas así etiquetadas no vuelven a cometer delito. En sentido contrario, los falsos negativos no perjudicarían, sino que beneficiarían, a personas blancas, pues la herramienta arroja un riesgo bajo de reincidencia en circunstancias que, con posterioridad, se verifica que finalmente aquellas personas reincidieron criminalmente.

Lo anterior es solo una muestra de los sesgos concretados, pero cabe preguntarse si es posible leer entre líneas el factor de raza o género en los datos que se utilizan para las predicciones. Pareciera ser legítimo sospechar de la información que se busca procesar,[24] a pesar de que expresamente, factores como la raza, deben excluirse siempre.[25] En este sentido, el problema transita desde lo exterior a lo interior, pues los resultados revelan el sesgo, pero en el cómo se producen está el núcleo problemático. La herramienta responde a una intención humana, y una equivocada conlleva un alto riesgo de que las herramientas desvirtúen por completo el objetivo central de una investigación, trasladando el foco desde los hechos presumiblemente delictuales y, por tanto, la determinación de culpabilidad o inocencia del sujeto, hacia los antecedentes personales del mismo, buscando el acoplamiento de estos a un perfil sospechoso, que

21. Carta Ética Europea sobre el uso de la IA en los sistemas judiciales y su entorno (2018: 7).

22. Mismo rol que le asigna incluso el proyecto de ley en nuestro país, materializado en el principio de diversidad, no discriminación y equidad.

23. Así se desprendió en el caso COMPAS. Véase el estudio elaborado por ProPública, disponible en https://tipg.link/S6X1.

24. Por ejemplo, en el caso de COMPAS: situación domiciliaria; si el sospechoso cree que la comisión de delitos se produce por falta de oportunidades laborales, entre otras. Formulario íntegro disponible en https://tipg.link/S6XG.

25. Reñiría con la decimocuarta enmienda, relativa a la «igualdad de protección».

se satisface con datos íntimos o sensibles del procesado.[26] Siguiendo a Hannah-Moffat y Avila (2023: 557), cuando los algoritmos base de estas herramientas son entrenados, utilizan datos recopilados de comunidades históricamente marginadas, insertando indicadores de riesgo que, reflejando más bien políticas selectivas, se alejan de un comportamiento criminal real. El problema de este escenario es, siguiendo a Starr (2014: 6), que el lenguaje científico «sanea» o «camufla» la discriminación.

El cuestionamiento anterior debe leerse, además, bajo riesgo de utilizar herramientas que, siendo creadas para una determinada etapa procesal, se pretendan incorporar ampliamente al proceso. Una utilización torcida del instrumento resultar muy peligrosa (Nieva, 2018: 71). Sin embargo, sobre la base de que cada etapa puede utilizar una herramienta predictiva de riesgo distinta (Quattrocolo, 2022: 150), el problema persistiría, pues dice relación con la determinación de decidir sobre factores no vinculados directamente con el delito. Para Starr (2014: 806), el argumento en contra sigue en pie, independiente de la etapa procesal: usar variables no relacionadas con la comisión del delito. Si el tránsito inter etapas opera sin sobresaltos, podría desatenderse el objetivo de cada etapa procesal y expandirse el sesgo discriminatorio a todo el *iter* procesal.

Aquí, la importancia del diseño y producción de la tecnología de *machine learning* se hace evidente, pues habrá que recurrir el interior técnico de la herramienta para identificar el origen de dichos sesgos. En efecto, la configuración de una *machine learning* consta de ocho pasos (Lehr y Ohm, 2017: 655), que son los siguientes: *problem definition, data collection, data cleaning, summary statistics review, data partitioning, model selection, model training, and model deploymen,* siendo algunos de ellos «pasados por alto» (Lehr y Ohm, 2017: 704) en la discusión sobre la producción de los sesgos discriminatorios y de sus posibles soluciones. Dicha configuración implica no pocas decisiones de programación, generando una legítima inquietud en torno a la fácil —pero a la vez sutil— posibilidad de perpetuar ciertas desigualdades traspasables al ejercicio jurisdiccional, cuando este no responde a la investigación de los hechos, sino más bien a la inserción del procesado en perfiles específicamente delimitados. Cómo lúcidamente plantea Nieva (2018: 74 y 75), «los datos desvinculados de la autoría delictiva no pueden ser tenidos en cuenta», razón por la cual se ha expresado (Solar, 2022: 60) que se requiere de una delimitación clara y previa de «factores o atributos jurídicamente protegidos» para colaborar anticipadamente con la garantía de imparcialidad, excluyendo estos factores de la base de datos con la que trabajará la herramienta.

Sobre el primer paso, la definición del problema no debe responder a intereses personales, sino que debe ser el resultado de un entendimiento inter-subjetivo sobre qué es lo que realmente se desea calcular, dado que «no todo es medible» (Lehr y Ohm,

26. En el presente trabajo, los términos «datos íntimos» y «datos sensibles» se utilizan indistintamente. Sobre esta última noción, parte de su contenido puede verse en nota 44.

2017: 673) a través de una IA basada en *machine learning*, pues presenta límites naturales de precisión que no son posibles de desatender (Kehl, 2017: 11-12). La recopilación de datos y la limpieza de los mismos resultan cruciales para definir correctamente de dónde se extraerán estos y cuáles serán reservados y eliminados para las fases posteriores de desarrollo.[27] En cuanto a la partición de datos, es bastante ilustradora a efectos de diferenciar los tipos de datos[28] que se tratan en el desarrollo en comento.

Por su parte, la demanda de elaboración o selección[29] de algoritmos o modelos que sean explicables o entendibles, responde a la selección del modelo, pues, si se pretende usar para el ejercicio jurisdiccional, se deberá optar por aquellos que prevalecen la capacidad de explicar u ofrecer razones para las predicciones (Lehr y Ohm, 692 y 693).

En conclusión, si la publicidad y la transparencia presentan sus propios opuestos, la imparcialidad del juzgador se ve soterrada con máquinas que perpetúan sesgos discriminatorios, cuyo origen es posible identificar en el proceso de diseño y producción de la herramienta, cobrando vida en la entrega de predicciones erradas.

La intimidad del procesado frente a la IA en materia criminal

Sorteando el largo debate doctrinario sobre las nociones de intimidad y privacidad, así como de su contenido,[30] en este apartado se explicarán las dimensiones de la intimidad que se abordarán, por qué importan al proceso criminal y, en consecuencia, a las herramientas predictivas de riesgo de reincidencia, y en último término, cómo es que constituye un problema la relación intimidad-IA para el despliegue del derecho de defensa en juicio.

27. Factores como la raza o el género no debiesen considerarse en los datos que ingresarán a la herramienta que efectúa la predicción, pues la determinación del riesgo no debiese responder a esos factores.

28. Existen los «datos de entrenamiento», los «datos de prueba» y los «datos del mundo real». Todos ellos responden a objetivos distintos en la producción de la *machine learning*. Siguiendo a Lehr y Ohm (2017: 684 y siguientes), la relevancia en la etapa de «partición de datos» es debido a que la evaluación del algoritmo que se desarrolla no se agota con los «datos de base» o de «entrenamiento», sino que, se debe medir el algoritmo con los «datos de prueba», que, a su vez, deben ser distintos a los «datos reales», pues estos aparecen solo cuando el algoritmo entra en funcionamiento. Ahí se produce el «primer contacto». Que el algoritmo funcione «correctamente» con datos de entrenamiento y prueba no garantiza una correcta representación de la realidad. Aparece con ello la importancia del «seleccionador» de los datos, que debe conocer el medio para saber «si los datos invisibles del mundo real serán similares a los datos recopilados» (Lehr y Ohm, 2017: 687 y 688).

29. Es posible crear el modelo algorítmico o bien guiarse por otros ya existentes (como los árboles de decisiones o las redes neuronales), según refieren Lehr y Ohm (2017: 688 y 689).

30. Bastará tener presente lo indicado por Toscano (2017: 535), quien, con base en la obra de Julie C. Innes, recuerda el «caos de la privacidad» que se revela al adentrarse conceptualmente en su significado.

En primer término, es un hecho que la intimidad se encuentra deteriorada,[31] expresándose, a consecuencia de su relación con el derecho de protección de datos personales, la amenaza que para ella significa el desarrollo de la IA, pues estos son su «alimento» (Nieva, 2018: 142; Bueno de Mata, 2020: 27). A pesar de ser derechos fundamentales distintos, la frontera que les separa es muy delgada, y es por ello que Kulhari (2018: 23 y 24) apunta que, si bien la Carta de Derechos Fundamentales de la UE los separa en dos artículos (7 y 8), esta diferencia sería más formal que de fondo, pues existe una vinculación natural entre ambos. Con todo, la inviolabilidad de las comunicaciones, la vida familiar o el «derecho a estar solo» *(right to be let alone)* forman parte de una noción histórica de privacidad, cuya protección jurídica ha sido desarrollada en el tiempo y, por tanto, no es reemplazable por la que entrega el derecho de protección de datos personales, dado que es imposible que comprenda todas las dimensiones indicadas (Kulhari, 2018: 23). Sobre lo dicho, es importante indicar que la aparición de la noción de «datos personales» sencillamente desafió las categorías tradicionales de privacidad, y que, en el caso de la UE, se trató de una categoría no contemplada al momento de redacción de la Carta de Derechos Fundamentales (Quatroccolo, 2022: 48).

Luego, con el desarrollo de una sociedad globalizada e hiperdigitalizada, se ha producido una simbiosis que ha decantado hoy, desde mi perspectiva, en una constante superposición del contenido histórico y de los datos,[32] cuya agrupación masiva o «big data» sirve para identificar tendencias y correlaciones generales con posibilidad de afectar directamente el comportamiento de los individuos,[33] permitiendo, como nunca antes, un procesamiento de información continua, siendo esto último el elemento clave (Alonso, 2022: 157) de la revolución tecnológica y, en consecuencia, un actor protagonista del desarrollo de la IA. Así, la «recolección indiscriminada de datos» que permite el «big data» (Nieva, 2018: 151), en un lenguaje atingente a este trabajo, no es otra cosa que la recolección indiscriminada de la intimidad.

Este deterioro responde a un doble fenómeno: uno centrado en el individuo y otro en las corporaciones. A pesar de existir las dimensiones clásicas,[34] la intimidad humana se refleja hoy en una cantidad enorme de datos,[35] dentro de los que habrá algunos muy sensibles y otros que quizás no forman parte del contenido esencial de la intimi-

31. Véase la acuciosa descripción de la recolección de datos que se puede efectuar por parte de distintas corporaciones privadas, en el contexto de lo que se denomina «capitalismo de la vigilancia», en Véliz (2021: 17-37).

32. El «petróleo del siglo XXI». Sobre este concepto, véase Hoffman-Riem (2019: 53-57).

33. Véase «Opinion 03/2016 on purpose limitation» del «Article 29 Data Protection Party», disponible en https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm.

34. Aunque incluso la noción de «comunicaciones» y «hogar» se ha cambiado sustancialmente en este tiempo. Sobre estos dos conceptos, véase Quatroccolo (2022: 58 y ss.).

35. Que para el «big data» comprende aquellos que están incluso más allá de los «datos personales», según Hoffman-Riem (2019: 52).

dad.[36] Lo anterior, además de ir difuminando la frontera intimidad-datos personales, provoca un diálogo con nuevas figuras como el «derecho al entorno virtual».[37] En la era digital, el ser humano es más transparente y renuncia a su intimidad, quizás sin la consciencia debida.

El segundo elemento deriva del comportamiento de las corporaciones privadas, que desarrollan su actividad en un mercado global de datos (economía de datos)[38] muy desregulado y sin controles eficaces (Pelayo, 2020: 63; Quatroccolo, 2022: 45), provocando que los datos de las personas lleguen a manos de distintas empresas (como Alphabet [Google], Amazon, Meta o Apple, algunas de las «big tech» más importantes en la actualidad), que los utilizarán con fines claramente lucrativos. Más allá de que la preocupación por la relación intimidad-datos personales e IA es un punto latente en los propios productores de la misma, el interés económico logra superponerse.[39] Lo expuesto en cuanto al deterioro es clara señal de la insuficiencia de la auto regulación, pues presenta una casi total responsabilidad de las corporaciones, en comparación a la del individuo que «auto expone» su intimidad. En este sentido, siguiendo a Edwards y Veale (2017: 66 y 80), el débil consentimiento en materia del tratamiento de datos personales en red no presenta siquiera «apariencia de autodeterminación informativa», y tan solo legitima la extracción de datos personales en favor de un tercero. Terceros que, en el caso de las corporaciones, presentan una auto regulación o coregulación con un «mal historial en materia de privacidad». ¿Podría entregarse, en consecuencia, el desarrollo de las herramientas a estas corporaciones?

El segundo punto es determinar por qué la intimidad importa al proceso criminal, y ello se debe a que constituye una piedra angular del proceso investigativo (Nieva, 2018: 150), encontrándose en tensión con la actividad de investigación criminal (Quatroccolo, 2022: 45; Esparza, 2021: 279), toda vez que esta última intentará extraer prueba que justamente integra el continente de intimidad del investigado. A pesar de que no todos los frutos de la investigación criminal sean datos personales (Esparza, 2021: 81), ellos configuran, como se dijo, una dimensión de la privacidad, ya sea en su vertiente de «privacidad de la información» (Leenes y De Coca, 2018: 282) o derechamente de la

36. El artículo 4 del Reglamento General de Datos Personales (UE, 2016) indica que los datos personales son «toda información sobre una persona física identificada o identificable», pudiendo ser una fotografía, el estado civil o el domicilio de una persona, entre otros elementos. Los datos personales sensibles son aquellos que revelan el origen racial o étnico, las opiniones políticas, las convicciones religiosas o filosóficas, datos relativos a la salud, entre otros.

37. Véase Bueno de Mata (2020: 8 y 9).

38. Al igual que Véliz (2021: 337), la economía de datos, en esta ocasión, la relaciono directamente con «datos personales», más allá de que podría haber una economía de «datos impersonales».

39. El 2023 marcó una clara tendencia negativa en sanciones a las empresas privadas por infracciones al Reglamento de Protección de Datos Personales de la UE. La estadística puede seguirse de manera actualizada en https://tipg.link/S6jH.

dimensión de los «datos personales de la privacidad» (Quattrocolo, 2022: 49), siendo en ese estado perfectamente procesables por la herramienta. Es ahí donde radica su potencialidad, pues muchos elementos pueden hoy transformarse fácilmente en un «dato».

En este esquema, la interacción entre estos dos derechos (intimidad y datos personales) resulta crucial a la hora del procesamiento de los datos, pues el segundo derecho referido es un instrumento valioso para que se precise de lo mínimo y necesario para el tratamiento lícito de los mismos, teniendo una función de «trazar límites» (Esparza, 2021: 282) frente a estos tratamientos. Empero, en el proceso criminal, esto parece fácilmente superable, pues el interés público en la persecución se sobrepone a la intimidad, limitando tal derecho con creces.

Un último punto se relaciona con la intimidad como elemento problemático. Dado que la intimidad se encuentra deteriorada, es fácil extraer de ella datos personales que podrían formar parte de los datos de entrenamiento o los datos de prueba, según las definiciones del desarrollador. Esto significa que puede existir un traslado de datos personales desde la esfera de intimidad de las personas a algunas de las etapas de diseño y producción de la herramienta (escenario pre-procesal), con todos los impactos que ello puede producir, especialmente en materia de sesgos (escenario procesal).

Si lo que estos instrumentos hacen, en base a la técnica de *machine learning*, es desarrollar la capacidad de correlacionar datos e identificar un perfil criminal de riesgo,[40] y junto con ello, indicar si el sujeto evaluado cumple con dicho perfil, clasificándole y asignándole una determinada puntuación, el problema es que dicho perfil podrá haber sido definido a partir de los datos de una masa de individuos, apartándose de la comisión del delito o la investigación criminal desarrollada. Ello conduce al riesgo de que, siendo lo que es (solo un perfil criminal que arroja una puntuación de reincidencia), represente para el proceso lo que realmente no es (un medio de prueba que colabore con la acreditación de la culpabilidad o inocencia de una persona que es sospechosa de delito).

Existirán, en consecuencia, distintos datos sensibles en el contexto del proceso criminal. Aquellos obtenidos con medidas intrusivas del ente persecutor, cuyo traslado de información (datos íntimos) desde la esfera de intimidad hasta el proceso criminal está legalmente justificado —pues ni la intimidad ni la protección de datos personales son derechos fundamentales absolutos—,[41] aunque subsista la posibilidad de

40. Que responderá a una teoría psico-criminológica. Quattrocolo (2022: 158) explica que, por ejemplo, la teoría que inspira a COMPAS correlaciona los antecedentes criminales del reo y las respuestas del cuestionario, por un lado, y el conjunto de datos de grupos sociales y étnicos, por otro, sirviendo estos últimos para evaluar el riesgo de comportamiento violento del individuo analizado.

41. Véase el reporte de la Agencia de la Unión Europea para las Derechos Fundamentales, titulado «Gettin the Future Right Artificial Intelligence and Fundamental Rights», página 61, disponible en https://tipg.link/S6wK.

impugnación tradicional en caso de actuación lesiva. Y aquellos datos que ingresarán a la herramienta, que correlacionando esos datos íntimos en base a una teoría psicocriminológica, arrojará una determinada predicción de reincidencia, conllevando el riesgo latente de completar una eventual insuficiencia probatoria (Nieva, 2018: 101), que sin destruir la presunción de inocencia, se suple con una predicción de riesgo, bajo el alero de ser un producto objetivo dotado de una suerte de presunción de veracidad, o más *ad hoc* al nuevo lenguaje, de una «presunción *a priori* de confiabilidad en los programas informáticos» (Quattrocolo, 2022: 167).

En conclusión, frente a una intimidad débil, la extracción de datos es sencilla para el desarrollador, y esos datos pueden destinarse a la elaboración de una herramienta predictiva que, habiendo superado las etapas de diseño de la herramienta, quedará apta para funcionar protegida legalmente, pero con reales posibilidades de hacerlo de manera opaca y sesgada, cerrando un descuidado circuito de datos que, con base en la intimidad de distintas personas, puede generar el juzgamiento al procesado, más que por lo que hizo en relación a un delito, por los caracteres que forman parte de su intimidad, pues el contenido de ella podría coincidir con un perfil de riesgo previamente configurado que asignará mayor o menor puntuación a los elementos de dicho contenido, desviándose por completo el objeto de un juicio criminal.

La configuración de la vulneración del derecho fundamental a la intimidad

Los desarrolladores de herramientas podrían vulnerar la intimidad, correspondiendo identificar el acto vulneratorio. Este puede producirse en contextos pre-procesales, a propósito del diseño y producción de la IA basada en machine learning, y en el propio proceso (juicio oral). Para la configuración del acto vulneratorio, se debe confrontar el derecho fundamental a la intimidad del procesado y su derecho fundamental a la protección de datos personales, frente a la herramienta de IA.

A estas alturas del proceso criminal se habrá sorteado la fase de instrucción, por lo que el objeto al cual responderá la herramienta predictiva, en su diálogo con la investigación, será reafirmar o contradecir la visión del tribunal, sea que este se halle, con la investigación y práctica probatoria, más cerca de la inocencia o de la culpabilidad del sospechoso de delito. Así, el escenario que se ha dibujado hasta aquí, si bien no centra la vulneración de la intimidad en aquellos casos de investigación o recogimiento de evidencia, nos conduce a distinguir entre los datos recopilados como antecedentes probatorios *per se*, y los antecedentes de la predicción del riesgo, pues ello ayuda a observar nítidamente el acto vulneratorio relacionado a los últimos.

El acto vulneratorio comenzaría y terminaría en la fase de recogida de datos, y el sujeto activo del mismo será la entidad desarrolladora de la herramienta. Dichos datos se encasillarán según su utilidad en valores determinados previamente por el desarrollador. Si se pone atención al Reglamento 679/2016 y la Directiva 680/2016 de

la UE,[42] ambos centrados en el tratamiento de datos personales, desde mi perspectiva, se podrían identificar una serie de condiciones que abren paso al acto vulneratorio contra la intimidad del sujeto procesado. Este acto vulneratorio podría materializarse en dos espacios. Como el Reglamento General de Protección de Datos (UE, 2016/679) se vincula al tratamiento de datos personales (más allá del proceso criminal), la vulneración de la intimidad del sujeto podría concretarse en un contexto no procesal y en un tiempo en que la persona no era un procesado o investigado, sino que simplemente presentaba patrones que eran atractivos al desarrollador, para la práctica en etapas de producción o diseño de la herramienta. Sobre el quebrantamiento de la esfera que protege sus datos y su intimidad, valga lo que se dijo específicamente a propósito del problema de los sesgos, la recolección indiscriminada de datos y el arrastre que puede generarse desde dicha etapa hasta aquella en que finalmente se procesan los datos reales. Esos sesgos discriminatorios se van desarrollando desde la recogida de datos, pudiendo reafirmarse en el tiempo. Sin embargo, este acto vulneratorio podría no importar al proceso criminal (la persona podría no llegar a ser objeto de investigación criminal, a pesar de haber servido como alimento para identificar patrones de comportamiento).

Cuestión distinta es aquella intimidad de la persona que ya está siendo procesada, pues la normativa que rige el escenario propuesto es la Directiva 680/2016.[43] A pesar de que entre sus motivaciones se deja entrever una protección del derecho de protección de datos personales, su contenido es muy ambiguo.[44] De partida, dicha normativa no habla en ninguna disposición de IA, de tal modo que no ha sido pensada para las potencialidades de esta tecnología. Solo es posible inferir su utilización al indicar entre sus líneas la existencia de los «tratamientos automatizados de datos». Ello configura ya un destiempo legislativo.

La Directiva define esta actividad como de alto riesgo, y con ocasión de ello pretende establecer límites, como la necesidad de un contrato con la entidad que efectuará el tratamiento (considerando 11); que este sea «lícito, leal y transparente» (considerando 26); la posibilidad de prescindir de él,[45] y el establecimiento de las «autoridades de control».[46] Sin embargo, se deja abierto igualmente el camino para que corporaciones privadas obren con libertad sobre la producción y el diseño de la IA, pudiendo materializarse los conflictos tratados previamente, como los sesgos discriminatorios (artículo

42. Relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos.

43. Artículo 9.1; Quattrocolo (2022: 57).

44. Una crítica exhaustiva de la Directiva puede encontrarse en Ruggieri (2021: 320-324).

45. Según el considerando 26 de la Directiva 680/2016 (UE): «Los datos personales solo deberían ser objeto de tratamiento si la finalidad del tratamiento no puede lograrse razonablemente por otros medios».

46. A pesar de que su competencia no comprende procesos jurisdiccionales, según lo dispone el artículo 45.2.

11.3)[47] y la opacidad normativa. Lo anterior, se explica por dos motivos. En primer lugar, la regulación final se entrega a cada Estado miembro (Quattrocolo, 2022: 56) y, en segundo lugar, a pesar de que existen derechos que en teoría permiten al procesado fijar límites al tratamiento, como se ha advertido, el equilibrio entre la privacidad y la investigación criminal no son nada fáciles (Quattrocolo, 2022: 56). Un «derecho a la revisión humana» o la prohibición de discriminación parecen no lograr contrarrestar el poder externalizado. Incluso, se ha dicho que, por más que se establezcan «aparentemente una serie de derechos de acceso» en los artículos 13, 14 y 16 de la Directiva, no hay garantía de acceso a los códigos algorítmicos mediante los cuales se realiza el tratamiento (Quatrocolo, 2022: 57).

Y es que si se aplicara el reglamento no existirían mayores diferencias: el acto vulneratorio podría perfectamente concretarse, y la intimidad del procesado verse derrotada por la recogida de datos. Dado que la Directiva no plantea más que un bosquejo general de protección, pues, se deja abierto el proceso a la regulación del Estado miembro,[48] se obtiene como resultado que toda una estructura normativa de tratamiento de datos personales con fines de índole criminal se entrega a las personas que conducen los Estados, pudiendo ellas confiar en terceros privados para el tratamiento.

En conclusión, una cosa es el conjunto de datos íntimos del procesado, allegados al proceso a raíz de medidas investigativas intrusivas, pero la incorporación de dichos datos en una herramienta predictiva del riesgo —para lo cual se podrá hacer uso de dichos datos, sea en su totalidad o parcialidad, lo que depende de cómo se bosquejó la herramienta y de lo que permitan los Estados— para que dicho riesgo sea una prueba del proceso, es otra cuestión muy distinta. Diferentes preguntas subyacen a esta extracción de intimidad: ¿por qué se permite que una herramienta predictiva acceda a dicha esfera jurídica, sin autorización judicial, extrayendo datos íntimos, en circunstancias

47. Dicho artículo señala: «La elaboración de perfiles que dé lugar a una discriminación de las personas físicas basándose en las categorías especiales de datos personales establecidas en el artículo 10 quedará prohibida, de conformidad con el Derecho de la Unión». A pesar de ello, el problema es doble. La forma de funcionamiento de la IA evidencia que, no queriendo discriminar, se pude producir el efecto de igual manera (me remito a lo dicho en el apartado «Sesgos frente a la imparcialidad del juzgador»). Y junto con ello, se pueden autorizar las decisiones, si el Estado miembro lo decide, yendo en contravención a la intención del artículo 11.1, que señala que: «Los Estados miembros dispondrán la prohibición de las decisiones basadas únicamente en un tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos negativos para el interesado o le afecten significativamente, salvo que estén autorizadas por el Derecho de la Unión o del Estado miembro a la que esté sujeto el responsable del tratamiento y que establezca medidas adecuadas para salvaguardar los derechos y libertades del interesado, al menos el derecho a obtener la intervención humana por parte del responsable del tratamiento».

48. El considerado 49 resulta particularmente relevante: «Cuando los datos personales sean tratados en el transcurso de una investigación penal o un procedimiento judicial en materia penal, el ejercicio de los derechos de información, acceso a los datos personales, rectificación o supresión de estos y la limitación de su tratamiento podrá ejercerse de conformidad con el Derecho procesal nacional».

de que ellos podrían eventualmente ya constar en el proceso criminal (recogida de evidencia)? ¿Quién podría garantizar que, si el funcionamiento algorítmico es opaco, tanto técnica como normativamente, esa intimidad no ha sido vulnerada con antelación al inicio del proceso criminal, en un contexto pre-procesal?

La vulneración del derecho fundamental a la intimidad del procesado se producirá cuando, para el funcionamiento de estas herramientas, que tendrán como objetivo arrojar un nivel de riesgo que sea utilizable por el juez —en especial en la sentencia—, estas se alimenten de un conjunto de datos íntimos, sin que el afectado pueda ejercer control en la extracción, en el posterior tratamiento y, en último término, en su incorporación al proceso. Esto último, como antecedentes del resultado del riesgo predicho, transformando en ilícito el tratamiento de los datos incorporados, desde su recogida hasta su traducción en una predicción de riesgo concreta.

La configuración de la vulneración del derecho fundamental de defensa

En este escenario, el derecho fundamental de defensa podría ser el último recurso para reconducir el objeto principal de la investigación criminal, sin embargo, su vulneración también podría fácilmente concretarse. Un primer punto del acto vulneratorio surgirá cuando se mantenga la opacidad normativa. Sin acceso al elemento clave de la IA, no se podría siquiera desplegar el contradictorio. Si bien se ha expresado la necesidad del lenguaje algorítmico, malamente podrá ser útil si, a pesar de adquirirlo la defensa, no se revela cómo funciona el algoritmo que tratará los datos íntimos, pues si se revelara, se podrían esgrimir argumentos en contra de los aspectos técnicos de la herramienta. Si lo legal encubre lo técnico, una condición mínima sería abrir el candado de la propiedad industrial. La publicidad del algoritmo sería la llave para abrir dicho candado y obtener información, restableciendo el estándar de garantía a la publicidad como elemento integrante del derecho de defensa, y fortaleciéndola frente a una propiedad industrial que aparece preliminarmente sobreprotegida.

Un segundo elemento del acto vulneratorio concurrirá cuando se incorpore al proceso la predicción. ¿Qué naturaleza jurídica tendría en el contexto de un juicio oral? Si se considera un medio de prueba, debiese responder al marco jurídico tradicional del derecho a la prueba, pudiendo impugnarse el resultado. Lamentablemente, este elemento del acto vulneratorio será deudor en gran medida del tratamiento de la opacidad normativa. Un tercer elemento del acto vulneratorio es la afectación del derecho a la motivación, el cual se vulnerará cuando el riesgo de «delegación» o «comodidad» en el ejercicio jurisdiccional del juez[49] se concrete al enfrentar el resultado de

49. Las voces de «comodidad» (Nieva, 2018: 75 y 104; 2022a: 100), «delegación» (Faggiani, 2022: 522) o «deresponsabilización» (Basile, 2019: 28-29) previenen sobre el riesgo de una actitud pasiva del juzgador, que permita entregar la labor jurisdiccional a la máquina, desplazándose, sin que sea posible —en un contexto de inaccesibilidad a la fórmula algorítmica— evidenciar ante el juez los errores de la herramienta, pues la defensa no tendría cómo hacerlo.

predicción, desprendiéndose del deber de motivación debido a una confianza ciega en el obrar de la herramienta, debilitándose igualmente el derecho a la prueba en su arista de valoración.

Un cuarto elemento afectado sería el derecho al recurso. Si el juez que dicta la resolución no motiva su sentencia, debido a la concreción del «efecto anclaje» (Christin y otros, 2015: 8) respecto del riesgo predicho, será casi imposible para la defensa impugnar adecuadamente la resolución. En uno de los precedentes judiciales sobre la materia, no se pudo asignar correctamente una función[50] a la predicción del riesgo. Lo anterior, valga la redundancia, conlleva el riesgo inherente de una motivación débil, que no se superpone ni colisiona con la motivación que el juez podría hacer del resto de los medios de prueba, sino que se centra en el juez frente a la predicción. La necesidad del lenguaje algorítmico sería indispensable también para los jueces, pues este les permite acercarse al entendimiento sobre la máquina y su función procesal. De no ser así, difícilmente se podrá controlar la práctica de esta prueba, y mucho menos su incorporación en la decisión jurisdiccional.

El impacto procesal de estas vulneraciones

Veámoslo progresivamente. Existe un proceso criminal abierto. La herramienta está legalmente incorporada a dicho proceso, pues ha sido entregado el desarrollo y funcionamiento a un desarrollador privado que ha configurado un perfil psico-criminológico validado por el Estado que le ha confiado esta gestión. Llegado el momento de procesar los datos íntimos reales, se vulnera la intimidad en la etapa de recolección, pues se permite la incorporación de distintos antecedentes que pueden ser de carácter estático o dinámico (responderán al perfil y teoría del desarrollador), y se adjuntan a otros que podrían ser parte del historial de antecedentes criminales del procesado o incluso obtenidos a través de la investigación criminal del ente acusador (en este último caso, la disponibilidad estará más que justificada). Ya efectuado el tratamiento, la máquina predecirá el riesgo de reincidir criminalmente.

Casi como si fuese aplicación de la adquisición procesal, el proceso recibirá esa predicción. La lógica indica que esto debiese operar después de la práctica de la evidencia probatoria. El examen del juez podría conducir a dos caminos. Se desconfía de la herramienta, apareciendo un rol activo del juzgador cuyo posible destino sea desechar o valorar negativamente la predicción de riesgo; o se confía en la máquina, pudiendo influir trascendentalmente en la decisión jurisdiccional de absolución o condena. En el tránsito entre la incorporación del riesgo predicho y la decisión del juez —que

50. Una contradicción que evidencian Kehl y otros (2017: 21) y Han-Wei y otros (2018: 10), pues no hubo explicación lógica del sitio que se debe asignar a los puntajes COMPAS. En este sentido, es posible hablar de una «influencia indeterminada en las decisiones jurisdiccionales» Kehl y otros (2017: 22 y 23).

debiese erigirse como un guardián de este micro-proceso de funcionamiento algorítmico— podría generarse una vulneración en cadena de los elementos del derecho fundamental de defensa, cuyo núcleo común y punto de partida sería la inaccesibilidad de la defensa a lo acontecido en dicho tránsito. Desde la recogida de datos íntimos hasta la sentencia, podría no tenerse herramienta alguna para interrumpir ese automatismo judicial, quedando cerrado un circuito que atentaría contra la intimidad y contra la defensa. En este escenario, habrá que pensar en condiciones de compatibilización.

Transparencia de origen y de funcionamiento

Los niveles de opacidad desarrollados al principio de este trabajo condicionaron el desarrollo del resto de los conflictos, siendo muestra del alto riesgo que representa para los derechos fundamentales tratados. Ante la colisión del derecho de propiedad industrial y del derecho a la intimidad y de defensa del procesado, no existe otro camino que abrir aquel candado primitivo de la opacidad, lo que provocaría la publicidad sin objeciones del algoritmo. Esta transparencia no se agota ahí, pues una vez que la herramienta funcione dentro del proceso, debe seguir sosteniéndose, a efectos de ir pesquisando las modificaciones que requiera durante su despliegue (Lehr y Ohm, 2017: 715). Dado que la técnica de machine learning no permite el descanso del aprendizaje de la herramienta, sería importante que ante el tratamiento de los antecedentes que haya perfilado el desarrollador, se autorice el acceso a los ajustes técnicos que pueda recibir la máquina, de lo contrario, nuevamente habrá intervenciones ocultas para el procesado y terceros, abriendo paso al secreto y a la legítima duda sobre la procedencia e imparcialidad de los datos.

Una transparencia de origen y de funcionamiento, que se imponga a la opacidad normativa, descomprimiría la tensión de origen abriendo paso a un debate distinto, enfocado en la opacidad técnica. En dicho debate, se tendrán que confrontar las técnicas de la herramienta. Si bien, la selección del modelo dependerá del objetivo de esta —si es que la toma de decisiones condujera al encarcelamiento, tendrán que prohibirse los «modelos particularmente inexplicables» (Lehr y Ohm, 2017: 657 y 715)— la apertura del algoritmo garantiza que se pueda dar lugar al contradictorio,[51] evitando la clausura

51. Incluso, la garantía de un «debate agonista» entre científicos de datos, abogados expertos y personas destinatarias del funcionamiento de la herramienta acerca de la implementación de esta, antes de su incorporación al proceso judicial, permitiría como acto reflejo herramientas o sistemas mucho más confiables y cuestionables (Hildebrant, 2018: 7). La idea del debate agonista dice relación con lo apuntado previamente, a propósito de la transformación del lenguaje desde el enfoque de los juristas (lenguaje algorítmico): es perentorio un diálogo interdisciplinario en la búsqueda de soluciones a propósito de la impronta de la IA en el proceso judicial. Sin embargo, garantizar el contradictorio una vez que se encuentre funcionando en el proceso judicial la herramienta, parece ser un mínimo irrenunciable para emplear la defensa procesal en caso de que dicho debate previo se frustre o no exista.

anticipada de la dialéctica del proceso judicial, animada por una confianza ciega en la tecnología, que se traduce en la asignación de un espacio sobreprotegido que allanaría el campo al automatismo jurisdiccional.

Auxiliándonos en lo que se ha propuesto en el seno de la UE, esta condición se consolidaría si se accede al algoritmo y al registro de los datos con los cuales se efectuó la práctica y entrenamiento de la IA, pues ello permite la identificación de los posibles sesgos en la selección del data set, que podrían reflejarse en el funcionamiento de la herramienta (Comisión Europea, 2020). Así, se podría intentar prevenir los riesgos que conlleva la transparencia sin sentido o «falacia de la transparencia» (Edwards y Veale, 2017: 21-23), debido a que se permitiría el acceso a una serie de informaciones relevantes para comprender el funcionamiento de la herramienta, alumbrándose la opacidad técnica, aun con las complejidades que ese proceso pueda implicar.

Esta condición por sí sola es insuficiente. La transparencia de origen, en consecuencia, debiese comprender al menos un acceso parcial al espacio pre-procesal, siempre que el conjunto de datos a evaluar se vincule claramente con la herramienta que sea utilizada en el proceso criminal y no responda a la lógica de recolección indiscriminada de datos. Esto permitiría a la defensa poder estar en mejores condiciones para efectuar un análisis técnico del funcionamiento de la herramienta, aunque no garantiza su comprensión. Sin embargo, es un punto de partida más ventajoso frente a la oscuridad de la caja negra. Además, reivindica otros beneficios, pues servirá también para efectuar procesos de auditorías, investigaciones científicas o evaluaciones mucho más informadas, cuestión que es imposible si se conserva y sobreprotege el secreto del algoritmo.

Delimitación del iter data o camino de los datos

Si las herramientas estudiadas requieren de la intimidad del procesado para su predicción, no es posible allanar el camino para que se traslade todo tipo de antecedentes a la máquina, por la sencilla razón de que se requieren para que funcione. De ser así, se validaría la desviación del objeto de la investigación, del proceso y del juicio oral, torciendo de entrada el uso de la IA.

Saber cómo se extraerán los antecedentes íntimos del procesado permitirá desde el inicio que la defensa sepa cómo enfrentarse al eventual acto vulneratorio, pues de lo contrario se asumiría una presunción de veracidad injustificada de licitud del tratamiento, en circunstancias que la línea que separa a los datos íntimos como antecedentes de prueba y como antecedentes del riesgo podría ser muy delgada. En dicho escenario, si la herramienta accediera a la intimidad, recibiendo los datos recopilados un tratamiento jurídico de evidencia probatoria que ha superado exámenes de admisibilidad y legalidad automática y anticipadamente, se entregaría un poder de vulneración al desarrollador sin precedentes.

Para evitarlo, una condición de compatibilidad es establecer la facultad de exclusión de datos íntimos, pudiendo ser ejercida por la defensa antes del tratamiento. Es, en el lenguaje técnico relacionado al *machine learning*, delimitar los datos reales. Vale prevenir que la intervención en este sentido no permite alterar nada de lo efectuado previamente en el proceso de diseño de la herramienta, por lo que contrarrestar los sesgos no será posible con la concurrencia de esta única facultad. Si se observa el tenor del Reglamento general de protección de datos y la Directiva 680, el funcionamiento de las herramientas predictivas coincide con la noción de «elaboración de perfiles» producto de tratamientos automatizados. A pesar de que en ambos se prohíbe, tan solo en uno de dichos instrumentos[52] se entrega una herramienta de defensa al interesado,[53] que es el derecho de oposición al tratamiento. Sin embargo, en el contexto del proceso criminal, dicha proyección es estéril, pues el derecho no se replica en la Directiva 680, que es el instrumento aplicable a la investigación criminal.[54] Así mismo, el despliegue de una oposición al tratamiento de datos íntimos parece improbable, pues su implementación procesal sería probablemente imperativa.

También resulta necesario determinar por qué extraer esos datos. Desde luego, aquí no parece haber otra solución que una definición conjunta (resultado del «debate agonista») de la necesidad de requerirse tales datos. En este caso, se deberá recurrir a teorías psico-criminológicas validadas que permitan definir cautelosamente una base de datos con lo suficiente para arrojar una predicción de reincidencia. Esta delimitación es importantísima, puesto que es previa al concurso de los datos reales, y establecería el esquema al que ellos ingresarían. La licitud de la recogida de los datos y el ejercicio de la facultad de exclusión dependerán de esta condición.

En último término, se tendrá que definir para qué calcular el riesgo en un contexto de juicio oral, donde la culpabilidad o inocencia del sospechoso tendrá como antecedente más próximo la evidencia obtenida de la investigación. Este último punto es el más delicado, pues una definición correcta reconduciría el objeto del proceso criminal (en caso de que este se desvirtúe), desde el foco del sujeto hacia el foco del delito. Esta parte del *iter* data, se relaciona con la necesidad de asignar un espacio procesal legítimo a las predicciones, estando estrechamente ligada a la definición previa del por qué extraer esos datos, pues dicha delimitación implica la asignación previa de una función procesal al puntaje de riesgo. Una definición conjunta (resultado del «debate agonista») vale también aquí para potenciar la condición, solo así se sabrá por qué importa al proceso la predicción de riesgo.

52. En el Reglamento general de protección de datos se encuentra en una disposición normativa (artículo 21, número 1), mientras que en la Directiva 680 tan solo es un considerando del reglamento (número 38).

53. El interesado, según el artículo 3.1. de la Directiva 680/2016 (UE), se trata de la persona física o identificable cuyos datos son objeto de tratamiento. En este caso, el involucrado en el proceso penal (artículo 6 letra a)).

54. Remisión expresa del Reglamento general de protección de datos a la Directiva 680/2016 (UE), en su considerando 19.

En conclusión, lo que aquí se propone es resguardar legalmente el camino de los datos (desde la recogida de datos hasta su traducción en puntaje de riesgo dentro del proceso), materializándose ello en el establecimiento de una herramienta concreta de exclusión de antecedentes íntimos antes de su ingreso a la máquina, cuando ellos no sean necesarios. La determinación de si son necesarios o no dependerá de los otros dos elementos del *iter* data, que son conocer la justificación de incorporación (por qué), vinculada simultáneamente con la función procesal (para qué) asignada a los resultados predictivos en el juicio oral.

En un escenario donde lo anterior no está definido, la relación entre la herramienta y el contexto procesal asignado perdería toda coherencia, debido a que transformaría en inexplicable e injustificable la utilización de la herramienta, fortaleciéndose una presunción de veracidad *a priori* sobre su uso, que atenta contra la intimidad y condiciona la defensa.

El derecho fundamental de defensa: Visualizar riesgos

El escenario al cual se enfrenta el procesado es adverso. Su intimidad está acechada por la investigación criminal, que buscará con medidas intrusivas penetrar su esfera de protección a efectos de recopilar prueba. Sumado a ello, ahora existe la posibilidad de que un software eventualmente opaco (normativa y técnicamente), en un ejercicio complejo de correlación de datos, entregue una predicción de reincidencia criminal que llega a manos del juez justo con ocasión de la audiencia de juicio oral, etapa previa a la dictación de sentencia.

La condición de compatibilidad que propongo en este escenario es previsualizar niveles de riesgo para que decidir una regulación u otra implique proyectar sus consecuencias jurídico-procesales. Un alto nivel de riesgo de vulneración conjuga dos elementos: la protección de la opacidad normativa y técnica, materializada en la prevalencia de la propiedad intelectual del desarrollador; y una desregulación del *iter* data, que implicaría no contar con la facultad de exclusión. El despliegue del derecho de defensa en este nivel sería casi imposible. Solo se podría contrarrestar el puntaje y, eventualmente, cuestionar la licitud en la obtención de los *inputs* o datos de entrada, ya no para excluirlos, sino para asociarlos a alguna categoría que podría estar vinculada a la protección de algún derecho fundamental distinto de los aquí estudiados (como la prohibición de discriminación),[55] y de esa manera, denunciar su vulneración y obtener una defensa procesal indirecta, pues el objeto a impugnar no sería expresamente ni la opacidad ni

55. Un camino no utilizado en el precedente de la materia, pero que da una señal de las opciones del despliegue de defensa en escenarios adversos de opacidad normativa y técnica. La denuncia de la violación de la enmienda relativa a la «igualdad de protección» habría concretado una mejor defensa en el caso *State of Wisconsin vs. Loomis*, pues hubiese ampliado el campo de pronunciamiento de la Corte Suprema de Wisconsin, según Kehl y otros (2017: 19).

la ausencia de la facultad de exclusión (a raíz de la desprotección del *iter* data), sino que la vulneración del derecho fundamental en la medida intrusiva o la sentencia.

Un nivel de riesgo medio será cuando alguno de estos dos elementos sea dispuesto en favor de la defensa. Si la opacidad normativa es liberada, la consideración de la predicción, como si se tratase del resultado de una prueba pericial,[56] daría pie a la posibilidad de impugnar técnicamente su funcionamiento, transformándose ello en el objeto principal a pesquisar. Aquí, el dominio del lenguaje algorítmico permitirá explicar al tribunal un mal funcionamiento del *software,* que podría deberse a que utiliza un modelo algorítmico errado o que se entrenó de manera incorrecta debido a una base de datos mal seleccionada,[57] generándose un sesgo discriminatorio en su predicción. En el segundo escenario, una herramienta de exclusión provocaría que al menos, a pesar de no ser público el algoritmo, se objete la incorporación de datos protegidos que, si bien serán correlacionados mediante algoritmo secreto, permiten discutir los datos reales que ingresarán al *software,* rescatándose una parte del *iter* data.

Un bajo nivel de riesgo operará cuando ambos elementos se regulen favorablemente, siendo público el algoritmo y además resguardando eficazmente el *iter* data. El acceso a una información completa abrirá paso a un amplio contradictorio, puesto que, correlacionando dicha información, podrá impugnarse fundadamente la predicción, existiendo la posibilidad de ejercer la facultad de exclusión y discutir técnicamente el funcionamiento. Sin dudas, podría defenderse la totalidad del *iter* data y además efectuar una defensa completamente técnica contra la herramienta, reduciendo la fuerza de su impacto procesal.

La autorización del Estado y una intervención humana amplia

Como colofón a estas propuestas, propongo una última condición. En primer lugar, los desarrolladores privados tendrían que estar autorizados por el Estado para que operen dentro del proceso criminal a través de la entrega del servicio de software en comento.[58] Ello no es más que un reflejo tangible de que el Poder Judicial es un poder del Estado, que resguarda con esta medida su independencia, imparcialidad y exclusividad del ejercicio jurisdiccional en materia criminal. A partir de un conocimiento previo y conjunto sobre la herramienta, que debiese ser fruto del diálogo[59] entre expertos de

56. Lo que implicaría someterla, por ejemplo, a los criterios Daubert. Véase De Miguel (2018: 48-51), que aborda críticamente dicho aspecto con ocasión del precedente judicial referido en la nota anterior.

57. Recordemos que una transparencia de origen y funcionamiento alcanzaría hasta el acceso a los registros de datos utilizados para la producción de la herramienta.

58. Lo indicado tiene su correlato en el RGPD (2016) y la Directiva 680/2016 (UE), pues estos instrumentos garantizan el derecho del «interesado» a no ser sometido a decisiones basadas únicamente en tratamiento automatizado, a pesar de que sea solo una intención general.

59. Como plantean Edwards y Veale (2017: 84), «cualquier intento de aprovechar los sistemas de apren-

IA y juristas, tendría que surgir naturalmente un modelo de autorización que permita tanto a los desarrolladores privados como a la defensa del procesado, respectivamente, seguir innovando y perfeccionando las herramientas predictivas, y desplegando el derecho fundamental de defensa. Será indispensable efectuar un examen de fiabilidad sobre la herramienta, pues su funcionamiento, sabemos, no es perfecto. Este modelo debiese habilitar al desarrollador para que su herramienta predictiva esté presente en el mercado con cuotas mínimas de información y garantías de funcionamiento.

El segundo elemento es totalmente indispensable y por tanto innegociable, al menos por ahora, y consiste en la presencia de una intervención humana amplia, pero previamente delimitada. El funcionamiento de la herramienta requerirá de supervisión humana, tanto para el ingreso de los datos como para el tratamiento con el algoritmo y el otorgamiento de los resultados.[60] Esta exigencia alude a una persona que ostente conocimientos técnicos suficientes para verificar si la herramienta está funcionando correctamente, pues ello garantizaría una regular optimización del sistema, que dependerá igualmente de la regulación de las tres condiciones anteriores. De esta manera, se resguarda el mantenimiento de la capacidad predictiva de la herramienta.

Conclusiones

La investigación efectuada permitió establecer la situación base sobre la que debe darse el debate que podría conducir a un eventual estado de compatibilidad entre los derechos fundamentales a la intimidad y de defensa del procesado, frente a la incorporación de IA en su modalidad de herramientas predictivas del riesgo de reincidencia, arrojando como resultado final una serie de posibles condiciones para alcanzar ese estado.

Para ello, se comenzó con el tratamiento de la opacidad, distinguiéndose sus niveles normativos y técnicos, así como los espacios de manifestación pre-procesales y procesales, apuntándose la posibilidad de que lo legal, encubriendo lo técnico, pudiese permitir un derecho de propiedad intelectual del desarrollador más fortalecido que el propio derecho de defensa. En cuanto a los sesgos, se identificó que el origen de ellos se asocia al espacio de diseño y producción de las herramientas basadas en la técnica de *machine learning* de la IA, y cobran vida durante el proceso, una vez incorporadas las predicciones de riesgo. En materia de intimidad del procesado, se logró acreditar que ante una intimidad débil, la extracción de datos es sencilla para el desarrollador, quien puede enfocarse en la elaboración de una herramienta predictiva que, habiendo superado etapas previas de diseño, quedará apta para funcionar con protección legal, con reales posibilidades de hacerlo de manera opaca y sesgada. Esto cierra un descui-

dizaje automático para el bien social, requiere de trabajos interdisciplinarios». La tarea mayor es ahora trabajar antes de la implementación.

60. Con una IA de «alto riesgo», la intervención humana es indispensable. Sin embargo, cuándo y qué cualidades técnicas debiese tener dicha persona, no es algo totalmente resuelto.

dado circuito de datos que, tomando la intimidad de distintas personas, puede hacer juicios basados en los caracteres que forman parte de la intimidad del procesado, pues el contenido de esta coincide con un perfil de riesgo previamente configurado, en vez de hacerlo basándose en lo que hizo la persona en relación a un delito.

Asimismo, se configuró una propuesta de vulneración del derecho fundamental a la intimidad del procesado, que se producirá cuando, para el funcionamiento de las herramientas, estas se alimenten de un conjunto de sus datos íntimos sin que el procesado haya podido ejercer en ninguna medida un control en la extracción de ellos, como tampoco en el posterior tratamiento, y en último término, en su incorporación al proceso penal, en su calidad de antecedentes del resultado del riesgo predicho por la herramienta. En el caso de la defensa, se producirá una vulneración de distintos elementos, cuyo inicio sería la inaccesibilidad algorítmica, que clausura el contradictorio. Un segundo elemento será la afección del derecho a la prueba, pues su naturaleza jurídica sería errática, a razón de la opacidad normativa que impediría la correcta objeción o impugnación. Luego, podría afectarse la motivación de la sentencia, debido a una posible actitud pasiva del juzgador frente a los resultados de la herramienta, y en último término, el derecho al recurso, debido a una posible motivación deficiente de la sentencia y una indeterminación de la función que se le asigne a la predicción de riesgo.

En materia de propuesta de condiciones de compatibilidad, se arribó a condiciones que en un funcionamiento integrado podrían permitir eventualmente una implementación que no vulnere la intimidad y el derecho de defensa. Las condiciones apuntan a una transparencia de origen y funcionamiento de la herramienta; una protección legal del camino de los datos, materializado en su punto de inicio en una facultad de exclusión; la previsualización de los riesgos que puede involucrar para el derecho fundamental de defensa una mala implementación de las herramientas, dividiéndose dichos riesgos en alto, medio y bajo; y en último término, la necesidad de autorización previa del Estado y la existencia de una intervención humana amplia, dotada de especialidad técnica en la materia.

Referencias

Amunátegui, Carlos (2020). *Arcana Technicae: El derecho y la inteligencia artificial*. Valencia, Tirant Lo Blanch.

Azuaje, Michelle (2023). «Propiedad intelectual como herramienta para promover la transparencia y prevenir la discriminación algorítmica». *Revista Chilena de Derecho y Tecnología*, 12: 1-34.

Article 29 Data Protection Working Party (2013). «Opinion 03/2013 on purpose limitation». Disponible en https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm

Basile, Fabio (2019). «Intelligenza artificiale e diritto penale: Quattro possibili percorsi di indagine». *Diritto Penale e Uomo.* Disponible en https://tipg.link/SAKJ.

Breiman, Leo (2001). «Statistical Modeling: The Two Cultures». *Satistical Science,* 16 (3): 199-231.

Bonta, James y Donald Arthur Andrews (2007). *Risk-need-responsivity model for offender assessment and treatment* (*User Report* No. 2007-06). Ottawa: Public Safety Canada.

Bueno de Mata, Federico (2020). «Macrodatos, inteligencia artificial y proceso: Luces y sombras». *Revista General de Derecho Procesal,* 51: 1-31.

Burrell, Jenna (2016). «How machines think: Understanding opacity in machine-learning algorithms». *Big Data & Society,* 3 (1): 1-12. Disponible en https://tipg.link/SAOO.

Carabantes, Manuel (2023). «Why artificial intelligence is not transparent: A critical analysis of its three opacity layers». En Simon Lindgren (editor), *Handbook of Critical Studies of Artificial Intelligence* (pp. 424-434). Cheltenham: Edward Elgar.

Chiao, Vincent (2019). «Fairness, accountability and transparency: Notes on algorithmic decision-making in criminal justice». *International Journal of Law in Context,* 15 (2): 126-139. Disponible en https://tipg.link/SAOv.

Christin, Angele, Alex Rosenblat y Danah Boyd (2015). «Courts and Predictive Algorithms». *Data & Civil Rights: A New Era Of Policing And Justice,* pp. 1-13. Disponible en https://tipg.link/SAPo.

Comisión Europea (2020). *Libro blanco sobre la inteligencia artificial: Un enfoque europeo orientado a la excelencia y la confianza* (pp. 23 y 24). Bruselas: Unión Europea. Disponible en https://tipg.link/SACy.

De Miguel, Iñigo (2018). «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling». *Law, Probability and Risk,* 17 (1): 45-53. Disponible en https://tipg.link/SAP4.

Dressel, Julia y Henry Farid (2018). «The accuracy, fairness, and limits of predicting recidivism». *Science Advances,* 4 (1). Disponible en https://tipg.link/SAP9.

Edwards, Lilian y Michael Veale (2017). «Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For». *Duke Law & Technology Review*, 16 (1): 18-84. Disponible en https://doi.org/10.2139/ssrn.2972855

Esparza, Iñaki (2021). «La Inteligencia Artificial y el derecho fundamental a la protección de datos de carácter personal». En Silvia Barona (editora), *Justicia algorítmica y neuroderecho: Una mirada multidisciplinar* (pp. 29-54). Valencia: Thompson Reuters Aranzadi.

—. (2022). «Derecho fundamental a la protección de datos de carácter personal en el ámbito jurisdiccional e inteligencia artificial. En especial la LO 7/2021, de protección de datos personales tratados para fines de prevención, detección, investigación, y enjuiciamiento de infracciones penales y ejecución de sanciones penales». En

Sonia Calaza y Mercedes Sánchez-Arjona (directoras), *Inteligencia artificial legal y Administración de Justicia* (pp.181-209). Pamplona: Thompson Reuters Aranzadi.

Faggiani, Valentina (2022). «El derecho a un proceso con todas las garantías ante los cambios de paradigma de la inteligencia artificial». *Revista Teoría y Realidad Constitucional*, 50: 517-546. Disponible en https://tipg.link/SAPz.

Garrett, Brandon L. y John Monahan (2020). «Judging risk». *California Law Review*, *108*: 439-493. Disponible en https://tipg.link/SAQO.

Gillespie, Tarleton (2014). «The Relevance Of Algorithms». En Pablo Boczkowski, Tarleton Gillespie y Kirsten A. Foot (editores), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-193). Cambridge: The MIT Press.

Hildebrant, Mireille (2018). «Algorithmic regulation and the rule of law». *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences,* 376 (2128). Disponible en https://tipg.link/SAQh.

Horcajo, Pedro, Víctor Dujo, José Manuel Andreu y Marta Marín (2019). «Valoración y gestión del riesgo de reincidencia delictiva en menores infractores: Una revisión de instrumentos». *Anuario de Psicología Jurídica*, 29 (1): 41-53. Disponible en https://tipg.link/SAgl.

Hoffman-Riem, Wolfang (2018). «Big Data: Desafíos también para el Derecho». Trad. por Eduardo Knörr Argote. Navarra: Civitas.

Kehl, Danielle, Priscilla Guo y Samuel Kessler (2017). «Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing». *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School*. Disponible en https://tipg.link/SAh4.

Kulhari, Shraddha (2018). *Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain Technology to Secure Privacy and Identity*. Múnich: Nomos. Disponible en https://tipg.link/SAhh.

Leenes, Ronald y Silvia De Conca (2018). «Artificial intelligence and privacy: AI enters the house through the Cloud». En Woodrod Barfield y Ugo Pagallo (editores), *Research handbook on the law of artificial intelligence* (pp. 280-306). Cheltenham: Edgard Elgar.

Lehr, David y Paul Ohm (2017). «Playing with the data: What legal scholars should learn about machine learning». *U.C. Davis Law Review*, 51 (2): 655-717. Disponible en https://bit.ly/3YlPMgb.

Liu, Han-Wei, Ching-Fu Lin y Yu-Jie Chen (2018). «Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability». *International Journal of Law and Information Technology,* 27 (2): 122-141.

Ludwig, Jeans y Sendhil Mullainathan (2021). «Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System». *Journal of Economic Perspectives*, 35 (4): 71-96.

Martínez, Lucía (2020). «Peligrosidad, algoritmos y *due process*: El caso State vs. Loomis». *Revista de Derecho Penal y Criminología,* 20: 485-502.

Miró, Fernando (2018). «Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots». *Revista de Derecho Penal y Criminología*, 20: 87-130. Disponible en https://tipg.link/SAk4.

Murphy, Kevin (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge: The Mit Press.

Nieva, Jordi (2018). *Inteligencia artificial y proceso judicial*. Madrid: Marcial Pons.

—. (2022a). «Un cambio generacional en el proceso judicial: La inteligencia artificial». En César Villegas Delgado y Pilar Martin-Ríos (editores), *El derecho en la encrucijada tecnológica: Estudios sobre derechos fundamentales, nuevas tecnologías e inteligencia artificial* (pp. 85-101). Valencia: Tirant Lo Blanch.

—. (2022b). «Inteligencia artificial y proceso judicial: Perspectivas ante un alto tecnológico en el camino». En Sonia Calaza y Mercedes Sánchez (directoras), *Inteligencia artificial legal y administración de justicia* (pp. 417-437). Pamplona: Thompson Reuters Aranzadi.

—. (2023). «Perder el control digital: ¿Hacia una distopía judicial?». *Revista de actualidad civil*, 4.

Pasquale, Frank (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Londres: Harvard University Press.

Pelayo, Ángel (2020). «Tic, inteligencia artificial y crisis de la democracia». En José Solar (editor), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho* (pp. 55-78). Madrid: Universidad de Alcalá.

Quattrocolo, Serena (2020). *Artificial Intelligence, Computational Modelling and Criminal Proceedings*. Cham: Springer.

Ruggeri, Stefano (2021). «Circulación de datos personales y tutela de derechos fundamentales en materia de justicia penal». En Silvia Barona, *Justicia algorítmica y neuroderecho: Una mirada multidisciplinar* (pp. 309-335). Valencia: Thompson Reuters Aranzadi.

Solar, José (2020). «Inteligencia artificial en la justicia penal: Los sistemas algorítmicos de evaluación de riesgos». En José Solar (editor), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho* (pp. 125-172). Madrid: Universidad de Alcalá.

Starr, Sonja (2014). «Evidence-Based Sentencing And The Scientific Rationalization of Discrimination». *Stanford Law Review*, 66 (4): 803-872. Disponible en https://tipg.link/SAkh.

Toscano, Manuel (2017). «Sobre el concepto de privacidad: La relación entre privacidad e intimidad». *Isegoría,* 57: 533-552.

Turner, Susan, James Hess, Charlotte Bradstreet, Steven Chapman y Amy Murphy (2013). *Development of the California Static Risk Assessment (CSRA): Recidivism*

Risk Prediction in the California Department of Corrections and Rehabilitation. Irvin: Universidad de California.

Véliz, Carissa (2021). «Privacidad es poder: Datos, vigilancia y libertad en la era digital». Trad. de Albino Santos Mosquera. Madrid: Debate.

Ward, Jeff (2021). «Black box artificial intelligence and the Rule of Law». *Law and Contemporany Problems,* 84 (3): 1-5.

Wexler, Rebecca (2018). «Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System». *Stanford Law Review,* 70: 1343-1429.

Sobre el autor

Manuel Urzúa Urzúa es abogado, licenciado en Ciencias Jurídicas y Sociales de la Universidad de Talca. Cursó el Programa de Magíster en Derecho con menciones de la Facultad de Ciencias Jurídicas y Sociales de la Universidad de Talca, obteniendo la mención en Derecho Procesal. Actualmente, es Profesor de Derecho de la Universidad de Talca, sede Talca. Su correo electrónico es manuelandresurzua@gmail.com.

REVISTA CHILENA DE DERECHO Y TECNOLOGÍA

La *Revista Chilena de Derecho y Tecnología* es una publicación académica semestral del Centro de Estudios en Derecho, Tecnología y Sociedad de la Facultad de Derecho de la Universidad de Chile, que tiene por objeto difundir en la comunidad jurídica los elementos necesarios para analizar y comprender los alcances y efectos que el desarrollo tecnológico y cultural han producido en la sociedad, especialmente su impacto en la ciencia jurídica.

DIRECTOR
Daniel Álvarez Valenzuela
(dalvarez@derecho.uchile.cl)

SITIO WEB
rchdt.uchile.cl

CORREO ELECTRÓNICO
rchdt@derecho.uchile.cl

La edición de textos, el diseño editorial
y la conversión a formatos electrónicos de este artículo
estuvieron a cargo de Tipográfica
(www.tipografica.io).