

XXXIII CONGRESO NACIONAL DE DERECHO PROCESAL

Mar del Plata, Argentina

14-15-16 de noviembre de 2024

“El Derecho Procesal a 30 años de la Reforma a la Constitución Nacional”

PENAL

Comisión

Tema 1: Inteligencia artificial: proceso penal y derechos fundamentales (arts. 75, inc. 19 y 22, CN).

Título

Inteligencia artificial y el riesgo de falsos positivos en el proceso penal

Autor

Román De Antoni

Abogado y docente de Derecho Procesal I de la Facultad de Ciencias Jurídicas y Sociales (Universidad Nacional de La Plata)

Mail: romandeantoni@hotmail.com

Síntesis de la propuesta

La ponencia tiene como objetivo abordar el impacto de la inteligencia artificial (IA) en el proceso penal, con un énfasis particular en los riesgos de error. Se examinan diversos casos en los que los algoritmos han generado falsos positivos y se analizan las posibles causas de estos problemas. Además, se exploran las tensiones potenciales con garantías constitucionales y se consideran propuestas regulatorias. El propósito principal de esta presentación es encontrar un equilibrio entre aprovechar los beneficios de la IA y proteger la integridad del sistema de justicia penal, asegurando el respeto al debido proceso.

Postulaciones

Premio de la Asociación Argentina de Derecho Procesal y Fundación de Estudios Superiores e Investigación (Fundesi)

Inteligencia artificial y el riesgo de falsos positivos en el proceso penal

“La inteligencia artificial nos obliga a revisar nuestra idea de justicia”.

David Weinberger

I. Introducción

El estudio del error judicial, especialmente el denominado “falso positivo” o condena de personas inocentes, ha adquirido una relevancia creciente tanto en el mundo anglosajón como en el europeo continental. Este interés ha impulsado numerosas investigaciones académicas y un activismo significativo orientado a la exoneración de personas injustamente encarceladas a nivel global¹.

Sus resultados han proporcionado una valiosa cantidad de información a punto tal de permitir la identificación de “prácticas probatorias” que, cuando se encuentran presentes en los procesos penales, aumentan la probabilidad de detenciones o condenas erróneas². No obstante, su foco se ha centrado principalmente en la “prueba tradicional”, es decir, en los errores derivados de la valoración de pruebas físicas o testimoniales. El problema es que, en los últimos años, la irrupción de programas de inteligencia artificial (IA), especialmente en las agencias de prevención y persecución del delito, ha hecho emerger un nuevo fenómeno que exige poner la lupa de fiscalización de los errores desde una nueva perspectiva.

¿Cuáles son los riesgos de implementar programas de IA en estas áreas? ¿Qué garantías constitucionales podrían verse afectadas? ¿Qué medidas deberían tomar los Estados para mitigar los errores? Estos son los interrogantes que se pretenden abordar en esta ponencia. Las inquietudes nacen como respuesta a corrientes que abogan por la aplicación urgente de

¹ El trabajo más conocido es el realizado por el *Innocence Project*, una institución creada en los Estados Unidos el año 1989 con el objeto de exonerar a personas erróneamente condenadas a través de la demostración de su inocencia utilizando preferentemente evidencia de ADN. El proyecto tiene oficinas en prácticamente todos los Estados norteamericanos e incluye oficinas internacionales en Australia, Canadá, Holanda, Irlanda, Nueva Zelanda, Reino Unido y, desde el año 2012, en nuestro país. Menos conocida, pero con mayor trabajo internacional, es el Innocence Network, que constituye una afiliación de 72 instituciones dedicadas a prestar servicios investigativos y de asistencia legal gratuitos para exonerar a personas condenadas erróneamente (Duce, M. Algunas Lecciones a partir de cuatro casos de condena de Inocentes en Chile. *Revista de Derecho* (Coquimbo), 22(1), <https://dx.doi.org/10.4067/S0718-97532015000100005>).

² En la literatura especializada, parece haber un acuerdo importante en que los principales factores que incidirían en su producción serían seis: (1) problemas con la identificación ocular de imputados por parte de víctimas y testigos; (2) uso de prueba pericial de baja confiabilidad y calidad; (3) uso de confesiones falsas; (4) uso de testigos mentirosos o poco confiables; (5) mal trabajo de las agencias de persecución penal; (6) inadecuada representación legal de la persona condenada (Duce, 2015).

salvaguardias para evitar que personas inocentes pierdan su libertad a causa de un error causado por un algoritmo.

II. Algoritmos en la prevención y persecución del delito: el riesgo de falsos positivos

Con el objetivo de aumentar la eficiencia de sus tareas, los Estados están integrando cada vez sistemas de inteligencia artificial en áreas claves como la seguridad nacional, la gestión fronteriza y la administración de la justicia. Este proceso tiene impactos notorios en el ámbito penal, donde se observa una tendencia al alza en la incorporación de programas basados en IA por parte de las agencias públicas encargadas de la prevención y persecución del delito.

En Europa y Estados Unidos, por ejemplo, podemos observar diversos softwares de reconocimiento facial que se utilizan para encontrar a personas investigadas o prófugas (SRF)³, así como programas que identifican disparos de arma de fuego a través de sensores auditivos colocados en la vía pública que dan alerta a las autoridades correspondientes⁴. Asimismo, existen sistemas de evaluación de riesgos de reincidencia delictiva⁵, programas

³ Existe una gran cantidad de softwares de reconocimiento facial que son utilizados por agencias de gubernamentales. Estos sistemas procuran encontrar, a través de técnicas de *machine learning*, coincidencias físicas entre la multiplicidad de puntos de los rostros que se pretenden evaluar con la finalidad de determinar si son idénticos o no. Así, por ejemplo, funciona el reconocido sistema de fotos de Google que se utiliza en dispositivos con sistema Android.

⁴ Uno de los softwares más conocidos es Shotspotter de la compañía SoundThinking. Este programa es utilizado para detectar disparos de arma de fuego a través de sensores auditivos colocados en la vía pública que detectan, a través de una triangulación acústica, la ubicación en donde se dispararon armas de fuego con el fin que las agencias policiales arriben inmediatamente y/o tener evidencias para dilucidar un caso. Si el sonido detectado y analizado por complejos sistemas de machine learning coincide con el de un disparo, el software envía una señal al Centro de Revisión de Incidentes. El tiempo total desde el disparo hasta la alerta es de menos de 60 segundos. Actualmente es utilizado en más de 150 ciudades en Estados Unidos y, en los últimos años, ha sido adquirido por algunos gobiernos de nuestra región como Uruguay y/o Paraguay.

⁵ Los programas de evaluación de riesgos son variados. En Estados Unidos, el más conocido es Compas de la compañía norteamericana Northpointe Inc. Es una herramienta que tiene como objeto predecir la probabilidad de que una persona pueda reincidir. Además, permite diferenciar el riesgo de que la persona cometa un delito ordinario o uno violento. El programa arroja un resultado probabilístico a partir del análisis de los datos suministrados. Entre los datos a ingresar (inputs) se destacan, por ejemplo, los resultados de las entrevistas personales y otros factores como la edad, antecedentes penales. El algoritmo sitúa al individuo evaluado dentro de un grupo que posee características similares y le otorga el puntaje de riesgo de ese grupo para que el órgano jurisdiccional tome la decisión. En otros países funcionan sistemas similares, por ejemplo, en Inglaterra funciona el programa OxREC. En España (Barcelona) se utiliza el sistema RESCANVI para los mismos fines y, desde el año 2007, la Policía Nacional utiliza el sistema VioGen, un sistema de predicción del riesgo en casos de violencia de género.

predictivos de delitos⁶, herramientas para detectar denuncias falsas⁷, sistemas de ciberpatrullaje⁸, softwares para el análisis de datos complejos en investigaciones criminales⁹, entre otros. No obstante, a medida que aumenta la adopción de estos programas, empiezan a trascender públicamente algunos casos de falsos positivos en la justicia penal, lo que genera serias preocupaciones acerca de la falibilidad de ciertos softwares que, de manera gradual, están comenzando a implementarse por distintas agencias gubernamentales de América Latina.

Los cuestionamientos son válidos. En Estados Unidos se han reportado numerosos casos de detenciones erróneas de personas inocentes debido al uso de sistemas de reconocimiento facial. Entre los afectados se destacan los casos de Robert William¹⁰, Porcha Woodruff¹¹, Michael Oliver¹², Alonso Sawyer¹³, Randall Reid¹⁴ y Nijeer Parks¹⁵. Otro caso notable es el de Michael Williams, un ciudadano 63 años que estuvo detenido por más de un

⁶ Con un objetivo similar al programa desarrollado por la agencia Precrime en la película *Minority Report* (2002), se han diseñado una serie de sistemas de IA destinados a promover la prevención de delitos. El país precursor de las técnicas de análisis predictivo es Estados Unidos, no obstante, desde hace más de una década estos programas se han empezado a adoptar también en la Unión Europea. Así, por ejemplo, encontramos Predpol (Estados Unidos), Anacrim (Francia), Keycrime (Italia), Crime Anticipation System (Países Bajos), Precobs (Alemania), Skala (España, Dinamarca y Austria). En América Latina en América latina, encontramos en Chile el sistema desarrollado por el Centro de Análisis y Modelamiento en Seguridad (CEAMOS) de la Universidad de Chile junto a Carabineros. El programa ha sido implementado en cincuenta y ocho comisarías, con el objetivo de dirigir de forma eficiente el patrullaje policial preventivo, definiendo áreas de mayor vigilancia y control.

⁷ En España encontramos el sistema informático "Veripol" diseñado por la Policía Nacional de España. El programa se encarga de detectar denuncias falsas en casos de robos con violencia e intimidación o tirones.

⁸ En algunos países se han implementado programas basados en big data para, a partir de su análisis, detectar delitos. Así, por ejemplo, encontramos en España el Maria Project, un Software que utiliza la Guardia Civil dedicado a la localización de plantaciones ilegales de marihuana mediante el tratamiento de datos de los consumidores de energía eléctrica que proporcionan las empresas. Por otra parte, con el objetivo de combatir delitos sexuales contra menores de edad, tales como el grooming o la pedofilia, existen programas que utilizan IA para crear perfiles falsos que, buscando ser atractivos para los agresores, suelen utilizarse como señuelos (también llamados "agentes provocadores"). Así, podemos encontrar, por ejemplo, el experimento de "Sweetie", un personaje virtual que creó la organización holandesa Terre des Hommes, que simula ser una niña filipina de diez años cuya imagen aparece en chats online y en páginas webs de citas. En pocos meses, los investigadores encontraron a veinte mil supuestos pedófilos que, además, intentaron pagar a la niña para mantener relaciones. Los datos de estos sujetos, pertenecientes a 71 países diferentes, fueron enviados a Interpol para intentar que las autoridades policiales y judiciales, intervinieran en la cuestión, lográndose a partir de ello algunas sentencias condenatorias.

⁹ En Europa se ha desarrollado un proyecto diseñado por la policía de West Midlands (Reino Unido) y de Amberes (Bélgica) -financiado por la Unión Europea- denominado VALCRI (Visual Analytics for Sensemaking in Criminal Intelligence Analysis). El sistema procesa datos y, a partir de una especie de razonamiento abductivo que produce el algoritmo, genera distintas hipótesis de investigación con el fin de orientar al investigador a lograr mayor agilidad y eficacia en su labor

¹⁰ Ver: <https://www.aclu.org/press-releases/man-wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart>, consultado el 5/6/2024.

¹¹ Ver: <https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html>, consultado el 20/4/2024.

¹² Ver: <https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/>, consultado el 06/3/2024.

¹³ Ver: <https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/alonzo-sawyer-facial-recognition-mistaken-arrest>, consultado el 6/3/2024.

¹⁴ Ver: <https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html?smid=url-share>, consultado el 17/4/2024.

¹⁵ Ver: <https://cnnespanol.cnn.com/2021/05/03/reconocimiento-facial-erroneo-envio-inocente-carcel-trax/>, consultado el 19/4/2024.

año acusado de un asesinato en primer grado por una alerta de Shotspotter¹⁶. Estos incidentes, que involucran principalmente a personas afroamericanas, han generado una gran preocupación en la sociedad civil y han dado lugar a acciones legales contra el uso de estos programas, incluyendo demandas colectivas en ciudades como Chicago¹⁷.

En Argentina, ha trascendido el caso de Guillermo Federico Ibarrola que, en marzo de 2019, fue detenido erróneamente en la terminal de Retiro producto de un match arrojado por el sistema de videovigilancia utilizado por el Gobierno de la Ciudad, que lo asoció a la búsqueda de un sospechoso de un robo agravado en Bahía Blanca. El sistema funcionaba reconociendo los rostros de los documentos cargados en el sistema de consulta nacional de rebeldías y capturas (CoNaRC). Una semana después, tras comprobarse que se trataba de una falla en la carga de la base de datos, recuperó su libertad. Tiempo después, tras una acción colectiva, la Cámara de Apelaciones de la Ciudad de Buenos Aires confirmó que el sistema, además de carecer de controles adecuados, afectaba el derecho a la privacidad de las personas¹⁸.

Otros sistemas de IA también han sido criticados por cometer errores debido a sus sesgos. Un ejemplo es el caso de *Viogen* en España, que ha clasificado erróneamente casos de violencia de género como "riesgo bajo"¹⁹. Otro es el programa policial *Veripol*, acusado de utilizar datos sesgados y de carecer de una evaluación rigurosa de las denuncias falsas²⁰. Asimismo, se ha cuestionado la falta de transparencia y la potencial discriminación en sistemas de policía predictiva, los cuales han sido criticados por generar

¹⁶ <https://www.chicagotribune.com/2021/08/20/how-ai-powered-tech-landed-a-chicago-grandfather-in-jail-for-nearly-a-year-with-scant-evidence/>, consultado el 16/4/2024.

¹⁷ Demanda contra el uso de Shotspotter disponible en: <https://www.macarthurjustice.org/wp-content/uploads/2022/07/Complaint-file-stamped.pdf>, consultada el 19/3/2024. En aquella presentación, se denunció que, durante un período de seis meses, la policía informó haber usado fuerza física contra al menos 82 habitantes de Chicago, casi todos hombres negros o latinos desarmados. La demanda citó específicamente varios casos de encarcelamientos erróneos a partir de alertas falsas de ShotSpotter. Con motivo de ello, en febrero de 2024, el nuevo alcalde de Chicago, Brandon Johnson, decidió no renovar la licencia de la compañía Ver: <https://www.theguardian.com/us-news/2024/feb/14/chicago-shotspotter-contract>, consultado el 11/3/2024.

¹⁸ La Sala I de la Cámara de Apelaciones de la Ciudad de Buenos Aires rechazó las apelaciones del Gobierno de la Ciudad y del Ministerio Público Fiscal y confirmó la sentencia de la jueza en lo Contencioso Administrativo de la Ciudad, Elena Liberatori, que en septiembre de 2022 señaló que el SRFP se implementó sin la debida protección de los derechos de los habitantes. Ver: <https://www.cels.org.ar/web/2023/04/confirman-la-inconstitucionalidad-del-uso-del-sistema-de-reconocimiento-facial/>, consultado el 15/4/2024.

¹⁹ Del Castillo C. (2022). Las víctimas denuncian fallos en VioGen, el algoritmo contra la violencia de género. Disponible. https://www.eldiario.es/tecnologia/victimas-denuncian-fallos-viogen-algoritmo-violencia-genero_1_8815201.html, consultado el 11/3/2024.

²⁰ Guerrero Cáceres, P. (2022). Algoritmos entrenados para ser racistas. Disponible en: <https://www.pikaramagazine.com/2022/11/algoritmos-entrenados-para-ser-racistas/>, consultado el 12/2/2024.

despliegues policiales desproporcionados en comunidades vulnerables y por carecer de evidencia empírica sobre su eficacia en comparación con métodos tradicionales²¹. De hecho, una investigación sobre 23.621 predicciones formuladas por el programa PredPol en Nueva Jersey reveló una tasa de acierto de "menos del medio por ciento", y un excesivo uso de estereotipos negativos²².

III. El riesgo de error: factores a considerar

Como podemos apreciar, las decisiones basadas en sistemas de IA no están exentas de errores. Los diseñados por seres humanos, pueden incorporar sesgos conscientes o inconscientes influenciados por las percepciones, prejuicios o limitaciones de sus creadores, afectando la calidad de los resultados. Aquellos desarrollados mediante técnicas de aprendizaje automático también pueden presentar fallas en su diseño y trabajar con datos inexactos, entre otros problemas. En función de ello, a continuación elaboramos un listado de causas frecuentes que contribuyen a la comisión de errores y que deben contemplarse por los operadores para disminuir la falibilidad al emplear este tipo de sistemas:

1. Falta de consideración de la naturaleza probabilística: Los algoritmos trabajan con probabilidades y no con certezas absolutas, lo que implica un margen de incertidumbre en sus resultados. Una interpretación incorrecta de este aspecto puede llevar a decisiones erróneas²³.

2. Falta de control de la precisión y exactitud en los datos: La calidad de los resultados de los programas de IA depende en gran medida de la calidad de los datos utilizados para entrenarlos y alimentarlos. Si los datos son incompletos, imprecisos o sesgados, los resultados serán poco confiables.

²¹ Sankin A, et. al (2021). *Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them*. Disponible: <https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them>, consultado el 10/3/2024.

²² Sankin A and Mattu S. Predictive Policing Software Terrible At Predicting Crimes. Disponible: <https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes>

²³ Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 36 (2020), párr. 31-36; y Panel de Alto Nivel sobre la Cooperación Digital, "The age of digital interdependence" (junio de 2019), págs. 17 y 18.

3. Expectativas irrealistas en la implementación: Las expectativas poco realistas de los sistemas de IA pueden llevar a la implementación de herramientas no preparadas para producir los objetivos deseados, lo que podría llevar a una sobreconfianza y una evaluación incorrecta de los resultados²⁴.

4. Sesgos discriminatorios y errores basados en prácticas de recolección sesgadas: Los algoritmos pueden reflejar discriminación racista o sexista debido a la inclusión de datos sesgados, lo que podría generar un rendimiento dispar en diferentes grupos demográfico²⁵. El uso de datos que no representan adecuadamente a toda la población puede agravar este problema. Un ejemplo común se encuentra en los sistemas de reconocimiento facial que, desarrollados mediante técnicas de *machine learning*, suelen ser entrenados con más fotografías de ciertos grupos, como hombres blancos, que de otros, como mujeres, personas negras y/o niños, lo que incrementa la probabilidad de falsos positivos en los grupos menos representados²⁶. Además, los datos históricos utilizados para entrenar modelos de IA pueden contener sesgos arraigados en prácticas pasadas. Por ejemplo, en Canadá, se descubrió que las personas negras sin historial delictivo tenían tres veces más probabilidades de ser arrestadas por posesión de marihuana que las personas blancas con un historial similar²⁷. Otros estudios empíricos han demostrado que los crímenes que afectan principalmente a las mujeres, como las agresiones domésticas y sexuales, estén subrepresentados en los modelos de vigilancia policial predictiva, ya que históricamente no se ha informado lo suficiente sobre ellos²⁸.

²⁴ Informe del Alto Comisionado de Derechos Humanos de Naciones Unidas. A/HRC/48/31. El derecho a la privacidad en la era digital. p 18.

²⁵ Ferrante, E. (2021). *Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos?* Revista Nueva Sociedad N°294. Disponible en: <https://biblat.unam.mx/es/revista/nueva-sociedad/articulo/inteligencia-artificial-y-sesgos-algoritmicos-por-que-deberian-importarnos>, consultado el 28/2/2024.

²⁶ A modo de ejemplo, una investigación del Instituto de Tecnología de Massachusetts reveló que el software de reconocimiento facial comercializado por Amazon identificaba erróneamente a las mujeres de piel más oscura como hombres el 31 por ciento de las veces Disponible: <https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html>, consultado el 25/3/2024.

²⁷ Rankin J., Contenta S. & Bailey A. (2017). Toronto Marijuana Arrests Reveal 'startling' Racial Divide", The Star. Disponible en <https://www.thestar.com/news/insight/2017/07/06/toronto-marijuana-arrests-reveal-startling-racial-divide.html>, consultado el 15/3/2024.

²⁸ Red Internacional de Organizaciones de Libertades Civiles. Aportes para el Informe de la ACNUDH: "El Derecho a la Privacidad en la Era Digital de Naciones Unidas". Consultado el 11/03/2024, <https://www.cels.org.ar/web/wp-content/uploads/2018/07/INCLO-OHCHR.pdf>, consultado el 14/3/2024.

5. Riesgos asociados al almacenamiento prolongado de datos personales: El almacenamiento prolongado de datos personales también plantea riesgos particulares, dado que estos podrían verse expuestos a nuevas formas de explotación que no se anticiparon al momento de su recolección inicial. Con el paso del tiempo, los datos pueden perder precisión, relevancia, o pueden incorporar errores de identificación históricos, lo que podría dar lugar a resultados sesgados o incorrectos en su tratamiento futuro²⁹.

6. Falta de transparencia en la toma de decisiones: Algunos sistemas de aprendizaje automáticos son “opacos”, es decir, desarrollan predicciones cuyo razonamiento subyacente es difícil o imposible de explicar, incluso para sus desarrolladores. Este es el problema que se conoce como “caja negra”³⁰.

7. Omisión de la influencia de actores en el diseño del algoritmo: Los resultados sesgados pueden provenir de la selectividad de actores que intervienen en el diseño del algoritmo. Esto se debe a que, en ocasiones, algunas voces relevantes no forman parte de la “mesa de conformación” del algoritmo. Esta ausencia puede aumentar el sesgo de selectividad en la recopilación de los datos, lo que conduce a obtener resultados direccionados, menos confiables y a aumentar prácticas discriminatorias³¹.

8. Problemas de interoperabilidad de sistemas: El riesgo de interoperabilidad en sistemas de IA se refiere a la dificultad que estos pueden tener para comunicarse con otras bases de datos de agencias públicas. Esto puede limitar la capacidad del sistema de IA para acceder a datos completos y actualizados de diferentes fuentes, lo que podría arrojar resultados erróneos o poco confiables³².

9. Falta de actualización continua: Los sistemas de IA requieren actualizaciones constantes tanto en sus algoritmos como en los datos que

²⁹ Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 36 (2020), párr. 33.

³⁰ Bleicher, A. (2017) Demystifying the Black Box That Is AI. Disponible: www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/

³¹ Callahan M. (2023). Algorithms Were Supposed to Reduce Bias in Criminal Justice—Do They?. Disponible: <https://www.bu.edu/articles/2023/do-algorithms-reduce-bias-in-criminal-justice/>, consultado el 19/3/2024.

³² Lezcano, J.M. (2023). La interoperabilidad en la administración de justicia bonaerense y la calidad del dato como elemento clave para el acceso a derechos. Disponible: <https://publicaciones.sadio.org.ar/index.php/EJS/article/view/507/419>, consultado el 8/5/2024.

utilizan para mantenerse precisos y confiables. La falta de actualizaciones puede llevar a que los algoritmos tomen decisiones basadas en información desactualizada o insuficiente, lo que puede resultar en errores significativos. Además, las leyes, regulaciones y contextos sociales cambian con el tiempo, por lo que los sistemas de IA deben adaptarse para reflejar estos cambios y evitar decisiones injustas o inexactas³³.

IV. El control de los errores: tensiones con derechos fundamentales

Los algoritmos operan a través de un código fuente que guía el análisis de datos según una fórmula diseñada por los programadores de las empresas desarrolladoras. Sin embargo, debido al carácter comercial de estos productos, los códigos fuente suelen mantenerse en secreto bajo la premisa del "secreto empresarial"³⁴.

En el ámbito penal, la revelación de los códigos fuentes se vuelve crucial para garantizar el derecho de defensa. ¿Qué sucede si se sospecha que un programa presenta errores, sesgos selectivos o trabaja con datos inexactos? ¿Tiene el imputado derecho a conocer el código fuente del software que afecta su caso?³⁵. Estos interrogantes nos ayudan a entender la magnitud de un problema que surge con la creciente incorporación de tecnología en los procesos judiciales.

Desde nuestra perspectiva, la falta de disponibilidad de los códigos fuentes, o su opacidad, no puede ser un argumento válido para sortear de un plumazo garantías fundamentales que, como explica Maier, han costado siglos construir³⁶. Esta imposibilidad de ejercer el control por parte de la defensa también conlleva la falta de transparencia respecto a cómo el Estado lleva a cabo sus tareas y fundamenta sus decisiones, impidiendo así verificar

³³ La Corte Suprema de Wisconsin advirtió este problema en el emblemático caso *State Wisconsin vs Loomis* donde precisó que “el programa COMPAS no explica cómo usa la información para calcular los puntajes de riesgo, la Guía para profesionales de COMPAS de 2015 de Northpointe, explica que las puntuaciones se basan en gran medida en información estática (antecedentes penales), con un uso limitado de algunas variables dinámicas”. Ver: <https://www.courts.ca.gov/documents/BTB24-2L-3.pdf>, consultado el 11/3/2024.

³⁴ Polanski, J. (2020). Garantías constitucionales del procedimiento penal en entorno digital. Buenos Aires: Hammurabi.p.101

³⁵ Roth, A. (2017) Machine testimony, en “Yale Law Journal”, vol. 126, n° 7, p. 1992.

³⁶ MAIER, Julio B. J., Derecho Procesal Penal, t. I, Pág. 763-766.

si existen errores potenciales que puedan causar daños y afectar derechos fundamentales.

Para abordar este problema, algunas corrientes han propuesto que los Estados deben encontrar mecanismos para generar códigos fuentes auditables e incentivar la innovación de forma tal que se respete el derecho de los imputados a conocer los algoritmos³⁷. Otras sugieren que las agencias gubernamentales establezcan una comisión científica para evaluar la confiabilidad de los algoritmos, siguiendo estándares como los que surgen de los conocidos casos “Frye”³⁸ o “Daubert”³⁹ de la Corte Federal de Estados Unidos⁴⁰.

Sin embargo, la solución más práctica parece residir en el ámbito legislativo. En este sentido, el Parlamento y el Consejo de Europa adoptaron en marzo de 2024 la primera regulación sobre el uso de IA. Esta iniciativa, conocida como “IA ACT”, buscar regular el uso de IA en función de su capacidad para causar daño, estableciendo reglas más estrictas para aquellas aplicaciones consideradas de “alto riesgo”. En tales casos, estipula que el “secreto comercial” no debe ser absoluto. Las empresas deberán revelar sus códigos fuentes cuando sean solicitados por las autoridades judiciales, quienes estarán obligados a cumplir con ciertos protocolos de confidencialidad⁴¹. Así, se entiende que la privatización del proceso no es absoluta, especialmente en el ámbito penal, donde el riesgo de error es demasiado elevado y el respeto por las garantías juega un rol fundamental.

5. Conclusiones

La irrupción de la inteligencia artificial nos conduce a una nueva era que merece poner mucha atención a los distintos programas que se aplican

³⁷ Kroll y otros (2017), accountable algorithms, en “University of Pennsylvania Law Review”, vol. 65; y Wexler (2018), Life, liberty and the secrets: intellectual property in the Criminal Justice System, en “Stanford Law Review”, vol. 70, 2018 (Citados en Polansky, 2020).

³⁸ Corte Suprema de Estados Unidos de América. *Frye v. United States*, 54 App.D.C.46, 293F. 1013, 34 A.L.R. (145), 1923. Allí, el alto tribunal estableció por primera vez el criterio de la “aceptación general en el área relevante” para la admisibilidad de pruebas periciales.

³⁹ Corte Suprema de Estados Unidos de América. *Caso Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579., 1993. Dicho tribunal estableció cuatro estándares mínimos a aplicar en la admisibilidad de la prueba pericial: primero, la refutabilidad (o científicidad); segundo, la publicación y/o revisión por pares; tercero, rango de error, y; cuarto, aceptación amplia y general de la comunidad experta.

⁴⁰ Polansky, J. (2020). *Ob cit.* p.101

⁴¹ Ver en <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206#footnote13>, consultado el 19/6/2024.

los poderes del Estado. La justicia penal no puede permanecer pasible a frente a los avances tecnológicos, pues hemos visto que sus fallas pueden derivar en enormes injusticias del sistema como es la detención o condena de personas inocentes.

Aunque las ventajas de utilizar algoritmos resulten evidentes, es importante tener en cuenta que los riesgos asociados a su uso suelen estar invisibilizados. De allí que, al implementarlos, es crucial considerar los potenciales errores y establecer requisitos que aborden aspectos fundamentales centrados en la calidad de los datos, la transparencia y la auditabilidad periódica de los mismos.

La imposibilidad de controlar los códigos es un asunto crucial que requiere poner especial atención en las futuras regulaciones para compatibilizar el respeto de las garantías fundamentales. El avance del fundamentalismo, consagrado en instrumentos internacionales de derechos humanos, así como el derecho de defensa, no admite que los estados utilicen pruebas de cargo obtenidas a través de fórmulas secretas. La protección del debido proceso es un pilar fundamental de cualquier sistema de justicia; de lo contrario, lejos de pretender avanzar con la implementación tecnológica, podríamos retroceder hacia la utilización de métodos irracionales basados en prueba de fe. ¿Cuál sería la diferencia entre condenar a alguien basándonos en pruebas de ordalías o en un algoritmo secreto? Aparentemente, ninguna.

En conclusión, en la era de la denominada “cuarta revolución industrial” la llegada de los algoritmos al proceso penal constituye un hecho inevitable. En esta ponencia hemos intentado advertir algunos de sus riesgos asociados a esta transformación para evitar que, a 30 años de la Reforma de nuestra Constitución Nacional, se perpetúen inequidades, discriminaciones o injusticias derivadas de un uso inadecuado o sesgado de la inteligencia artificial. Ello bajo la creencia que la preocupación por la reducción de errores no solo implica un imperativo de cualquier sistema de justicia, sino también un compromiso de los estados con la ética, el debido proceso y la protección continua de los derechos humanos fundamentales. -