

Inteligencia artificial y proceso penal: una cuestión de principios

En este artículo el autor realiza una aproximación inicial al establecimiento de principios para el respeto de los derechos fundamentales y las garantías constitucionales en la aplicación de la inteligencia artificial al proceso penal. Realiza un repaso de las posibles aplicaciones de la inteligencia artificial al proceso penal, describe los riesgos que conlleva el binomio proceso penal–inteligencia artificial para los derechos fundamentales y las garantías procesales y esquematiza una serie de principios rectores indispensables para compatibilizar inteligencia artificial, proceso penal y derechos fundamentales.



POR MARCO TERRANOVA

Abogado por la Universidad de Mendoza. Diplomado en Derecho 4,0 por la Universidad Austral. Maestrando en Derecho Penal en la Universidad Autónoma de Madrid y diplomando en Compliance, Ética Empresarial y Dirección de Procesos de Integridad en la Universidad de Mendoza.

I- Revolución 4.0

La Cuarta Revolución Industrial se diferencia de las tres anteriores por las siguientes razones: a) velocidad: la evolución se da a ritmos exponenciales, cada nueva tecnología engendra, a su vez, tecnología más nueva y poderosa. Todas las predicciones parecen erradas y todos los pronósticos acaban siendo superados al cabo de meses. El fenómeno de la “singularidad tecnológica” no parece ya una frontera sólo alcanzable para la literatura de ciencia ficción; b) amplitud y profundidad: no existe área o ámbito de incumbencia humano que no se encuentra alcanzado por esta revolución; c) impacto en los sistemas: se trata de la transformación de sistemas complejos entre (y dentro de) los países, las empresas, las industrias y la sociedad en su conjunto (Schwab, 2016).

Es posible identificar dos fenómenos disruptivos derivados de la revolución 4.0. El primero de ellos, la mutación radical de las nociones de espacio y tiempo a partir del uso masivo de nuevas tecnologías de la información y de la comunicación. El segundo, el surgimiento de nuevas formas de procesar los datos y la información en muchas actividades que antes sólo podían ser realizadas por nuestros cerebros (Corvalán, 2019). Inteligencia artificial (IA) –en sus diferentes variante–, *blockchain*, *big data*, internet de las cosas (IoT, por sus siglas en inglés), *cloud computing*, robótica, son sólo algunos de los principales protagonistas de esta transformación. Otros, como la computación cuántica, asumirán ese rol en el corto y mediano plazo, convirtiendo a los avances del presente en tecnología del pasado reciente.

II- Inteligencia artificial y proceso penal

El derecho, como toda actividad humana, no se encuentra exento de ser alcanzado por la revolución digital, máxime cuando la principal tarea de sus operadores es justamente la de procesar datos e información en clave jurídica. Una de las tantas oportunidades que se presenta es la posibilidad de incorporar las tecnologías disruptivas a la noble tarea de impartir justicia. En ese sentido, aquella que ofrece mayores posibilidades de aplicación, así como también peligros, retos y desafíos, es la inteligencia artificial (Nieva Fenoll, 2018).

No existe una definición universalmente aceptada de IA, en parte derivado de nuestra dificultad para definir a la inteligencia como capacidad cognitiva de los seres humanos. El glosario de la Carta Europea sobre el Uso Ético de Inteligencia Artificial en los Sistemas Judiciales y su Entorno de la Comisión Europea para la Eficacia de la Justicia (CEPEJ) del Consejo de Europa define a la IA como: “Un conjunto de métodos, teorías y técnicas científicas cuyo objetivo es reproducir, mediante una máquina, las habilidades cognitivas de los seres humanos”. (CEPEJ, 2019). A su vez, el núcleo de la IA son los algoritmos, que pueden definirse como “un conjunto preciso de instrucciones o reglas, o como una serie metódica de pasos que pueden utilizarse para hacer cálculos, resolver problemas y tomar decisiones”. (Corvalán, 2019, p. 32)

¿Puede una máquina resolver un planteo de prescripción?, ¿determinar la existencia de peligro de fuga?, ¿resolver acerca de la inocencia o culpabilidad de una persona?, ¿individualizar y determinar la pena aplicable a un sujeto declarado penalmente responsable? Si bien no con el alcance que perfilan los interrogantes planteados, en la actualidad existen numerosos ejemplos de IA aplicada al proceso penal. Quizás el caso más conocido y discutido sea “COMPAS” (*Correctional Offender Management Profiler for Alternative Sanctions*)¹, una herramienta de inteligencia artificial desarrollada por la empresa norteamericana Northpointe Inc., que realiza una valoración de riesgo (*risk assessment*) de reincidencia a través del perfilamiento del sujeto estudiado. La valoración de riesgo la realiza mediante la toma de diversos datos (nivel de estudios del examinado, índice de criminalidad de la zona de residencia, antecedentes condenatorios, consumos problemáticos, relación de proximidad con amigos, antecedentes penales de la familia y amigos, etc.)² que luego son analizados por el algoritmo para dar una triple predicción: riesgo de reiteración delictiva durante el transcurso del proceso (“*pre-trial recidivism*”), riesgo de reincidencia en general y riesgo de reincidencia en delitos violentos.

En Argentina existen algunos casos de IA orientadas al derecho penal, tales como la “calculadora de la prescripción” creada por el Instituto de Neurociencias y Derecho de la Fundación INECO, que cuenta con una *ratio* de acierto del 92%. Para arribar a este número se plantearon

1 En el siguiente enlace puede encontrarse una breve descripción de su funcionamiento y un documento con las Frecuently Asked Questions: http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf, consultado el 19/11/2020.

2 En el siguiente enlace se encuentra el cuestionario a través del cual COMPAS arma un “perfil” de la persona examinada: <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>, consultado el 19/11/2020.

cuatro casos de prescripción de dificultad intermedia y alta a estudiantes de posgrado en derecho penal de la UBA y otras universidades argentinas, quienes demoraron en promedio 20 minutos en resolverlos. Frente a cada uno de los cuatros casos debían responder la siguiente pregunta: ¿se encuentra prescripta la acción penal? La respuesta debía ser cerrada, por sí o no. Luego, los mismos casos se sometían a análisis de la calculadora de la prescripción, que demoraba cerca de un minuto en resolverlos. En un primer análisis, se demostró que en 75% de las respuestas de la máquina coincidían con aquellas dadas por los estudiantes. Las discrepancias se redujeron del 25% al 8% cuando a los estudiantes que habían resuelto en forma distinta a la IA se les daba a conocer los resultados de la máquina y se les pedía que revisen su decisión. Así, un 17% reconocieron su error y modificaron la respuesta. Finalmente, solo el 8% de los estudiantes mantuvieron su respuesta luego de ser confrontados con las respuestas de la IA. No obstante, ese reducido porcentaje de estudiantes justificó su respuesta en una interpretación jurisprudencial y doctrinal de una causal de suspensión que, de momento, es francamente minoritaria. Por ende, tranquilamente podría argumentarse que la máquina acertó en todos los casos. En este sentido, es por demás interesante el estudio para demostrar como las distintas interpretaciones jurídicas sobre una misma cuestión pueden jugar un rol decisivo al momento de configurar el algoritmo. Por el momento, la calculadora de la prescripción se encuentra limitada al ámbito investigativo y sin aplicación práctica (Haissiner y Pastor, 2019, p. 73-77).

Tampoco puede dejar de nombrarse a PROMETEA, una herramienta de IA desarrollada en el ámbito del Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires³, que posee numerosas funciones que van desde la menos compleja automatización (por ej., un sistema que cuenta plazos procesales) hasta los sistemas más sofisticados que utilizan aprendizaje automático para reconocer patrones y emitir una decisión o predicción jurídica sobre la base de criterios estadísticos. Algunas de estas funciones han sido probadas en el fuero penal contravencional y de faltas con muy buenos resultados: el operador que mensualmente generaba 172 documentos, a partir de la utilización de PROMETEA en conjunto con el sistema de gestión digital del juzgado pasó a generar 6800 documentos en el mismo periodo temporal (Corvalán, 2018). Si bien PROMETEA se encuentran en etapa de implementación en varias jurisdicciones provinciales –entre ellas en la Suprema Corte de Justicia de la Provincia de Mendoza y en el fuero tributario de la Primera Circunscripción de la Provincia de Mendoza–, su aplicación más promisoriosa se ha dado en la Corte Constitucional de la República de Colombia, donde se creó una herramienta específicamente adaptada a las necesidades del tribunal superior: PretorIA (Estévez, Fillotrani y Linares Lajarraga, 2020).

Los beneficios que conlleva la aplicación de IA a la administración de justicia no se dan sólo en materia de productividad, optimización de recursos, eficiencia y celeridad, sino que también inciden favorablemente en la calidad del proceso y, más importante aún, en la calidad

³ <https://www.ialab.com.ar>

de la decisión a la que se arriba concluido el mismo. La IA puede relevar al juzgador de la realización de tareas repetitivas, mecánicas y rutinarias (no por ello exentas de errores, sino todo lo contrario), permitiéndole concentrarse únicamente en resolver aquellas cuestiones de alta complejidad cuyo tratamiento requiere de mayor dedicación y esfuerzo, así como también de habilidades y capacidades cognitivas que de momento son propias del ser humano, como, por ejemplo., la capacidad de empatizar. A su vez, la máquina, en principio, se encuentra exenta de heurísticos y sesgos cognitivos, todo lo cual permite arribar a una decisión más racional, desprovista de prejuicios e intuicionismos, que muchas veces deciden la suerte del pleito, pese a que no se expresan en la sentencia. O que, aun expresándose, resultan incommovibles, como sucede con la inmediación y el juicio de credibilidad del testigo por su lenguaje gestual, su postura o actitud nerviosa, todos elementos que la psicología del testimonio ha demostrado que resultan poco menos que inútiles a efectos de valorar la credibilidad del deponente. Ni hablar del avance en materia de seguridad jurídica: el algoritmo puede ser absolutamente predecible⁴ e incorruptible (esto último, en el sentido “humano” de la palabra).

III- Riesgos de aplicación de la IA al proceso penal

No obstante, no debe mostrarse a la IA como la panacea y la solución última a todos los dilemas de la administración de justicia penal. Así como los puntos positivos son numerosos, también hay grandes desafíos y riesgos en la implementación de la IA al sistema judicial y al proceso penal, algunos de ellos relacionados directamente con los derechos fundamentales y garantías constitucionales de los ciudadanos. De esto ha dado cuenta la Comisión Europea, al señalar expresamente que el uso de esta tecnología puede provocar la conculcación de los derechos fundamentales a la tutela judicial efectiva y a un juicio justo (Comisión Europea, 2020).

Surgen diversas y variadas problemáticas con relacionadas con:

1. El principio de juez natural

¿Sistema decisor o *recomendador*? Suponiendo que adoptásemos un sistema decisor no debe perderse de vista que, aunque la máquina resuelva en el caso concreto, en última instancia (por el momento) detrás de una decisión siempre habrá un grupo de programadores que llevaron al algoritmo por un camino determinado, ¿debemos saber quiénes son?, ¿cómo sería su proceso de elección?, ¿podrían funcionar en una estructura privada, o necesariamente deben ser funcionarios públicos?

2. Las reglas de competencia territorial

Recordemos lo señalado en relación con la mutación radical de la noción de espacio, ¿tendrán sentido en un futuro las normas de competencia territorial cuando sea una máquina la que resuelva?

⁴ La predictibilidad de la IA depende de la técnica que utilice y su trazabilidad. Por ejemplo, aquellas que utilizan redes neuronales artificiales no son trazables y, por ende, tampoco puede predecirse ni explicarse del todo la decisión que finalmente tome la máquina.

3. La trazabilidad de la decisión a la que arriba el algoritmo y su incidencia en el derecho de defensa.

Si no puedo conocer cómo el algoritmo llegó a determinado resultado, me resultará imposible defenderme del mismo. La violación a las garantías y derechos fundamentales puede darse de dos formas. Sucede que algunos desarrollos de IA no son trazables, como, por ejemplo, las redes neuronales artificiales. A este tipo de algoritmos se los denomina “cajas negras” (*black boxes*). En otras palabras, no puede explicarse del todo por qué el algoritmo llegó a determinado resultado. Pero también se da la situación de que muchos de los desarrollos de IA trazables son producto de iniciativas privadas que, amparadas en el secreto empresarial, son muy reacias a develar el código fuente de los algoritmos. Tal como señala Polansky (2019, p. 143), esto es lo que sucede en Estados Unidos, donde gracias a la interpretación jurisprudencial que ha privilegiado el secreto empresarial por sobre la necesidad de revelación del código fuente de herramientas de uso forense, las corporaciones prefieren no registrar ni patentar sus algoritmos y mantenerlos en secreto bajo ese privilegio. Esta problemática se planteó por primera vez con COMPAS en el caso “State v. Loomis” resuelto por la Suprema Corte de Justicia del Estado de Wisconsin en 2016.

Eric Loomis se declaró culpable de dos delitos: intento de evasión de un oficial de tráfico y manejo de un automóvil sin consentimiento del propietario. COMPAS otorgó como resultado que existía una alta probabilidad de que Loomis cometiera nuevos delitos, particularmente de tipo violento. El juez tuvo en consideración la información brindada por la IA a efectos de determinación la pena aplicable y lo condenó a la pena de 6 años de prisión. El condenado recurrió la decisión hasta llegar a la Suprema Corte del Estado de Wisconsin. Uno de sus agravios era que se había violado su derecho al debido proceso, puesto que no se le había permitido conocer cómo funcionaba el algoritmo, por lo que le resultaba imposible desafiar su resultado. Si bien la Suprema Corte reconoció que COMPAS no fue creada para ser aplicada a la determinación judicial de la pena y desalentó su uso en futuros casos, consideró que si era utilizada con conocimiento de sus limitaciones y advertencias (en concreto: que podía llevar a discriminar a grupos minoritarios, que tomaba datos de reincidencia nacionales y no estatales, que al utilizar un método actuarial no permitía una adecuada individualización del resultado, entre otros), la consideración de su resultado no violaba el derecho al debido proceso. La utilización adecuada, según el tribunal, también requería que la decisión no se fundamentara únicamente en la información aportada por el algoritmo secreto. Sobre estas consideraciones, rechazó el recurso de Loomis⁵.

4. El principio de doble conforme

Si la decisión es tomada por una IA: ¿quién revisa lo resuelto por la máquina?, ¿sobre qué debe recaer la revisión?, es decir, ¿cuáles pueden ser los motivos de agravio? Una postura más radical podría incluso llegar a plantear si es necesario revisar la decisión del algoritmo.

⁵ Para un análisis más amplio de la sentencia se recomienda: KEHL, Danielle; GUO, Priscilla; KESLER, Samuel. (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, págs 18-21. Disponible en: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>

5. El principio de no discriminación

Antes mencioné que la máquina, *en principio*, se encuentra exenta de heurísticos y sesgos cognitivos. Lo cierto es que siendo la IA una creación humana es inevitable que los sesgos de sus programadores no se trasladen al algoritmo, los cuales, en ese caso, se ven aumentados exponencialmente. El mayor peligro reside en que los datos (*data sets*) que toma el algoritmo para arribar a una decisión ya se encuentren sesgados de antemano, como sucede muchas veces con las estadísticas de reincidencia o reiteración delictiva. Se encuentra sumamente estudiado por la criminología que el sistema de persecución penal es selectivo y que esta selectividad produce un efecto de “etiquetamiento” (teoría del *labeling approach*), perpetuando y agravando la criminalización de ciertos grupos sociales, como por ej., las clases socioeconómicamente desfavorecidas. Si esta información se traslada al algoritmo sin tomar medidas correctoras (por ej., otorgándole un peso menor frente a otras variables), la decisión de la IA puede encontrarse sesgada, otorgando como resultado una sistemática valoración de riesgo alta a personas que viven en localidades pobres o marginadas⁶.

No es que estas prácticas no se encuentren ya arraigadas en el sistema penal actual, sino que el peligro reside en que se vean reforzadas con la aplicación de los algoritmos o, incluso, justificadas, en tanto siempre resulta un argumento efectista apelar a la capacidad de procesamiento de datos de la IA, infinitamente superior a la humana.

6. El principio de responsabilidad penal por el hecho.

Relacionado con el punto anterior, una incorrecta configuración algorítmica puede hacernos caer en un derecho penal de autor. Especial atención merecen las técnicas de “perfilamiento”, cuya utilización puede servir de base a una “re-revitalización”⁷ de la noción de la peligrosidad.

7. El principio de aplicación de la ley penal más benigna.

Haissiner y Pastor (2019, p. 89) plantean el interrogante relacionado a un cambio de configuración algorítmica posterior al hecho, siendo más benigno que el “vigente” al momento de la comisión del hecho, ¿cuál debe aplicarse?

Estos son sólo algunos de los problemas que pueden surgir en la implementación de IA al proceso penal y cuyo abordaje no se pretende agotar en este trabajo. La tarea por emprender

6 COMPAS fue cuestionada por la ONG ProPublica por discriminación racial en perjuicio de la población afroamericana pobre: “How We Analyzed the COMPAS Recidivism Algorithm”, disponible en inglés en <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>, consultado en fecha 23/11/2020.

7 Se habla de “re-revitalización” puesto que la peligrosidad se encuentra en un momento de auge a nivel legislativo, tal y como demuestra la introducción de la libertad vigilada al Código Penal Español mediante la Ley Orgánica 5/10. El halo de cientificidad y perfección que rodea a la inteligencia artificial puede generar que su aplicación a los pronósticos de peligrosidad o valoraciones de riesgo sobre las cuales se asienta la adopción de medidas de seguridad opere como un impulso añadido de proporciones nada desdeñables, tanto a su ampliación normativa como a su justificación jurisprudencial.

es otra: plantear una aproximación inicial a una serie de principios rectores que podríamos utilizar de guía o marco orientador a efectos de movernos dentro de un ámbito mínimo de seguridad en materia de derechos fundamentales, sistemas de administración de justicia penal y algoritmos.

IV- Algunos principios rectores para una aplicación constitucional y convencional de la inteligencia artificial al proceso penal

Los diferentes usos y funciones que puede otorgársele a la inteligencia artificial dentro de las etapas del proceso penal (instrucción penal preparatoria, etapa intermedia, juicio oral y ejecución de la pena) hacen inconveniente dictar una regulación pormenorizada con vocación de generalidad para todos ellos, siendo conveniente desde mi óptica la aprobación de reglamentación específica que regule su implementación de acuerdo con las necesidades y exigencias propias de cada acto y momento procesal. Por el contrario, sí resulta imprescindible normativizar ciertos principios rectores que permitan establecer un estándar mínimo para todos los casos posibles de aplicación de la IA en la justicia penal.

El orden expositivo de los principios no sigue criterios de prelación ni temporales, sino que todos revisten de igual importancia y deben observarse en forma continuada durante los procesos de creación, desarrollo, implementación, adaptación y utilización de herramientas de inteligencia artificial a la administración de justicia penal. En otras palabras, y a riesgo de adelantarme al desarrollo de los principios, si bien el punto de partida es que toda aplicación de IA al proceso penal lo sea bajo el control total de un operador judicial humano, de poco sirve la que decisión final sea de una persona de carne y hueso si lo hace guiada por un resultado obtenido a través de un algoritmo secreto, sesgado, con defectos de diseño y susceptible de sufrir ataques por parte de agentes externos.

a) Principio de control y decisión final humana

El derecho internacional de los derechos humanos exige la presencia de personas investidas de un cargo público y con determinados atributos para entender en las causas judiciales (arts. 6.1 Convención Europea de Derechos Humanos y Libertades Públicas, 8.1 Convención Americana de Derechos Humanos y 14.1 Pacto Internacional de los Derechos Civiles y Políticos). La responsabilidad de tomar una decisión sobre los derechos de otro semejante, máxime cuando se encuentran en juego bienes jurídicos inalienables, tales como la libertad, debe recaer siempre sobre seres humanos.

Debe adoptarse el paradigma de “sistema recomendador”. El uso de la inteligencia artificial en el proceso penal adquiere un enfoque complementario (adoptar los beneficios de la inteligencia artificial sin resignar aquellos derivados de la actividad cognitiva humana y viceversa) y de colaboración con el operador jurídico, quien en cualquier caso podrá apartarse de la recomendación dada por el sistema de inteligencia artificial, aunque su utilización sea obligatoria. En otras palabras, si bien el funcionario puede estar constreñido a utilizar la inteligencia artificial, siempre podrá apartarse de su resultado. Sin embargo, la introducción de la inteligencia artificial al proceso penal en ningún caso debería relevar al funcionario

interviniente de la obligación de fundar sus resoluciones de acuerdo con lo dispuesto en la normativa procesal.

En el dictado de resoluciones de mero trámite y ordenatorias del proceso, donde la actividad jurisdiccional es mínima y la tarea queda delegada normalmente a funcionarios inferiores de la oficina judicial, la injerencia de la IA en el proceso decisonal –normalmente mecánico y de bajo riesgo en cuanto a sus consecuencias– seguramente será mayor. Respecto de las resoluciones que permiten pasar de una etapa del proceso a otra (por ej., clausura de la investigación preparatoria o control de la acusación), el dictado de medidas cautelares o de la propia sentencia, se espera una participación mucho más activa del funcionario judicial responsable. De cualquier modo, ello quedará sujeto a la reglamentación específica, pero lo que sí debe evitarse en cualquier supuesto es que la decisión final sea tomada por la IA o que la tesis adoptada por esta resulte vinculante y que el funcionario no deba motivar sus decisiones con el nivel de exigencia requerido por el derecho adjetivo.

De esta manera, además, se respeta el principio de juez natural, puesto que, si el juzgador de la causa estuviera obligado por lo resuelto por la inteligencia artificial, la decisión no sería suya, sino de aquellos que programaron la misma, quienes no gozan del carácter de juez natural de la causa. Por último, también se simplifica la problemática derivada de la revisión y revocación de la decisión tomada por la inteligencia artificial.

b) Principio de no discriminación

Toda aplicación de inteligencia artificial al proceso penal debe cumplir con el principio de no discriminación. Debe asegurarse que la configuración algorítmica no lleve a la máquina a tomar decisiones basadas en las características personales de los interesados que no tengan vinculación perceptible y justificada con el hecho objeto del proceso.⁸ Especial cuidado debe tenerse con los sistemas de “perfilamiento” y aquellos utilizados para realizar evaluaciones de riesgo (por ej., para decidir acerca de la existencia de riesgo procesal), susceptibles de derivar en un derecho penal de autor.

La Carta Europea sobre el Uso Ético de la Inteligencia Artificial en los Sistemas Judiciales y su Entorno hace especial referencia a este peligro al decir que: “Incluso si no están diseñados específicamente para ser discriminatorios, el uso de estadísticas e inteligencia artificial en los procesos penales ha mostrado el riesgo de provocar el resurgimiento de las doctrinas deterministas en detrimento de las doctrinas de individualización de la sanción” (2019, p. 66). Es por ello que el documento incluye dentro del apartado de “usos a ser considerados con las más extremas reservas” a los algoritmos aplicados al perfilamiento de individuos con el objeto de decidir sobre sus derechos en causas penales.

⁸ En ciertos casos las características personales deben ser tenidas en cuenta sin que ello implique necesariamente incurrir en una práctica discriminatoria, como por ej., en los casos de violencia contra la mujer, donde para decidir sobre la procedencia o no de una medida cautelar de carácter personal, el género de la víctima y del agresor deben ser tenidos en cuenta necesariamente, puesto que resultan elementos objetivos del tipo penal aplicable.

La misma preocupación surge del Libro Blanco de IA de la Comisión Europea: “Puede suceder que el uso de determinados algoritmos de la IA para predecir la reincidencia delictiva dé lugar a prejuicios raciales o de género, y prevea una probabilidad de reincidencia distinta para hombres y mujeres o para nacionales y extranjeros. Fuente: Tolan S., Miron M., Gomez E. and Castillo C. “Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia”, Best Paper Award, International Conference on AI and Law, 2019.” (2020, p. 14).

Como regla, debe evitarse que el algoritmo se encuentre configurado de manera tal que tome decisiones basadas en categorías sospechosas o, en terminología del Tribunal Constitucional español, diferenciaciones “odiosas”, motivadas en la raza, género, orientación sexual, opinión política, origen, nacionalidad, condición económica o cualquier otro rasgo o condición personal innata u opción elemental que expresa el ejercicio de las libertades más básicas (STC 62/2008, FJ. 5). Cuando alguna de estas categorías deba ser tenida en cuenta (por ej., nacionalidad a fin de determinar arraigo), nunca podrá ser la única variable a través del cual la máquina arriba a una recomendación. A su vez, debe asegurarse que el “peso” (*weight*) o relevancia dada a la misma dentro de la configuración algorítmica no la coloque en una situación de privilegio frente a las demás variables.

c) Principios de explicabilidad, trazabilidad, publicidad y máximo acceso algorítmico

Se deben adoptar los principios de explicabilidad, trazabilidad, publicidad y máximo acceso algorítmico a fin de asegurar la inviolabilidad de la defensa en juicio y la garantía del debido proceso. Desde una faz procesal estos principios tienden a asegurar la igualdad o paridad de armas, permitiendo al interesado desafiar la validez científica del algoritmo y, en consecuencia, la solución a la que este arriba.

La razón de ser de estos principios parte de un axioma procesal en materia de acusación y derecho a una defensa real y efectiva: resulta imposible cuestionar o impugnar lo que no se conoce. El proceso de toma de decisiones del algoritmo debe ser explicable y trazable en su totalidad. El resultado al que arriba el sistema de inteligencia artificial siempre debe poder ser explicado en cada uno de sus pasos en forma inteligible a las personas interesadas, especialmente al acusado. Se sigue en este sentido el criterio adoptado por las Directrices Éticas para una Inteligencia Artificial Fiable del Grupo independiente de expertos de alto nivel sobre IA constituido por la Comisión Europea (2019, p. 16): la utilización de “cajas negras” queda absolutamente prohibida en materia penal, donde el contexto y gravedad de las consecuencias derivadas del resultado al que arribe el sistema de inteligencia artificial justifican el máximo grado de explicabilidad.

Por otra parte, los principios de publicidad y máximo acceso algorítmico tienden a hacer efectivos los principios de explicabilidad y trazabilidad. De nada sirve que un sistema de inteligencia artificial sea explicable y trazable si la información no se encuentra a disposición de los interesados en tiempo y forma. El alcance de la obligación de publicidad debe ser lo más amplio posible, abarcando los conjuntos de datos y los procesos que dan lugar a la decisión

del sistema de inteligencia artificial, incluidos los relativos a la recopilación y etiquetado de los datos, así como también a los algoritmos utilizados. A su vez, deben comunicarse las finalidades y limitaciones del sistema, incluyendo, en su caso, información acerca del valor predictivo, nivel de precisión, sensibilidad y especificidad del algoritmo.

En cuanto a la forma, deben asegurarse los medios que permitan un acceso rápido y simple a la información en cuestión. Sin embargo, la publicidad de los datos objeto de procesamiento en un caso determinado, así como los resultados a los que arribe la inteligencia artificial en cada caso, quedan sujetos al régimen de publicidad específico previsto en el ordenamiento procesal y en las leyes de protección de datos personales. Por último, la publicación de los datos debe ser oportuna, estando al alcance de los interesados previo a su utilización dentro del proceso.

Para los casos en que la IA es aplicada a la obtención de evidencia (por ej., programas utilizados para la extracción forense de información alojada en dispositivos de almacenamiento electrónico de datos, programas de análisis genético automatizado, herramientas de medición del nivel de alcohol en sangre por aire aspirado, etc.), existe doctrina que niega la utilidad y necesidad de consagrar los principios de publicidad y máximo acceso algorítmico. Por el contrario, considera que debe ponderarse en el caso concreto la necesidad de revelar el código fuente, según si resulta o no indispensable para poder refutar el resultado al que arribe el algoritmo. Refutación que básicamente recae sobre determinar si la técnica aplicada cumple con los llamados criterios “Daubert” de la Suprema Corte de Justicia de los Estados Unidos. En otras palabras, si la confiabilidad del algoritmo puede establecerse en forma suficiente sin necesidad de conocer su código fuente (por ej., por su uso extendido y aceptado en otras jurisdicciones, por el consenso de la comunidad científica acerca de sus resultados, por la existencia de evidencia complementaria que robustece los resultados otorgados por la máquina), prevalece el secreto empresarial, sin que exista afectación alguna al debido proceso (Spolansky, 2020).

Más allá de la posibilidad de aceptar esta excepción a la aplicación generalizada a los principios de máxima transparencia algorítmica y publicidad, es necesario realizar algunas aclaraciones: a) se debería mantener en toda su vigencia la regla de prohibición de “cajas negras”. Supongamos que en el caso concreto se llega a la conclusión de que es absolutamente necesario conocer el código fuente para determinar si el resultado es confiable o no, pero, una vez obtenido, su conocimiento resulta inútil puesto que nos encontramos con que el algoritmo no es trazable ni explicable; b) más allá de su accesibilidad, deben prohibirse o, por lo menos utilizarse con las más extremas reservas, las técnicas de perfilamiento a través de IA; c) sólo debería reconocerse la excepción para las técnicas dirigidas a la obtención de evidencia mediante la utilización de IA, nunca a su aplicación directa a la emisión de decisiones dentro del proceso penal (por ej., un algoritmo que, en base a los hechos del proceso y la legislación aplicable, dictamine sobre la extinción de la acción penal por prescripción o la aplicación de determinada magnitud de pena).

d) Principios de seguridad y fiabilidad

Finalmente, se requiere la consagración de los principios de seguridad y fiabilidad a fin de asegurar la garantía a una tutela judicial efectiva, adoptando un enfoque de prevención y minimización de los posibles daños generados por la utilización de la inteligencia artificial en el proceso penal. Si bien la decisión final queda en manos del operador jurídico humano, no puede desconocerse que la recomendación o información otorgada por el sistema de inteligencia artificial incidirá en ella. Esta incidencia puede ser crucial, sobre todo cuando se trate de cuestiones de menor complejidad, donde el operador jurídico se sienta confiado en apoyarse acríticamente en lo recomendado por la máquina. Tampoco puede mostrarse a la inteligencia artificial como infalible, sino reconocer que, al igual que los seres humanos, es susceptible de presentar equivocaciones que pueden llevar a resultados erróneos⁹. Una vez reconocida dicha circunstancia, deben adoptarse las medidas tendientes a minimizar los riesgos.

Previo a la implementación de los sistemas de inteligencia artificial debe asegurarse que los mismos sean seguros y fiables. En cuanto al primer requisito, deben adoptarse todas las medidas que dentro del estado de la técnica contribuyan a proteger al sistema de inteligencia artificial frente a las vulnerabilidades que puedan permitir su explotación por parte de agentes externos malintencionados. Por su parte, la fiabilidad se refiere a la construcción y funcionamiento del sistema en sí mismo, el cual, previo a su implementación, debe haber superado en forma satisfactoria procesos de verificación, validación y evaluación adecuados. El contenido específico de los mismos deberá protocolizarse para cada caso y según los conocimientos de las técnicas y métodos científicos aplicables. La información del sistema cuya publicidad pueda poner en peligro su integridad y hacerlo vulnerable podrá ser secreta en la medida de lo estrictamente necesario, sin que la reserva pueda conllevar una afectación de los derechos de defensa en juicio y acceso a una tutela judicial efectiva. No así en lo relativo a su fiabilidad, cuya información se encuentra plenamente alcanzada por la obligación de publicidad y máxima transparencia algorítmica.

Referencias bibliográficas

Comisión Europea. «Directrices Éticas para una IA fiable». Bruselas: 29 de noviembre de 2019. <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.

Comisión Europea. «European ethical Charter on the use of Artificial Intelligence in judicial

⁹ Sobre los distintos defectos que puede presentar una herramienta de IA aplicada al proceso penal: ROTH, Andrea (2017). Machine Testimony. *Yale Law Journal* (126:1972), 1989-2000. Disponible en: <https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=5810&context=yjlj>, consultado el 30/11/2020.

- systems and their environment». Estrasburgo: 2018. <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.
- Comisión Europea: «Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza». Bruselas: febrero de 2020. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>.
- Corvalán, Juan Gustavo. «Hacia una Administración Pública 4.0: digital y basada en inteligencia artificial. Decreto de “Tramitación digital completa”. Revista Jurídica Argentina La Ley, n.º 2018 D (agosto de 2018): 917-36.
- Estévez, Elsa, Fillotrani, Pablo, y Linares Lajarraga, Sebastián. «Prometea». Washington: Banco Interamericano de Desarrollo, junio de 2020. <https://publications.iadb.org/es/prometea-transformando-la-administracion-de-justicia-con-herramientas-de-inteligencia-artificial>.
- Giles, Martin. «Qué ha logrado Google con su supremacía cuántica y qué no». MIT Technology Review en español, 26 de septiembre de 2019. <https://www.technologyreview.es/s/11493/que-ha-logrado-realmente-google-con-su-supremacia-cuantica-y-que-no>.
- Haissiner, Martín D. y Pastor, Daniel R. Neurociencias, tecnologías disruptivas y tribunales digitales. 1.ª ed. Buenos Aires, Argentina: Hammurabi, 2019.
- Kehl, Danielle, Guo, Prisculla, y Kesler, Samuel. «Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing». Responsive Communities Initiative. Cambridge, Massachusetts: Berkman Klein Center for Internet & Society, Harvard Law School, 2017. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>.
- Larson, Jeff, Kirchner, Lauren, y Angwin, Julia. «How We Analyzed the COMPAS Recidivism Algorithm». ProPublica, 23 de mayo de 2016. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.
- Nieva Fenoll, Jordi. Inteligencia artificial y proceso judicial. 1.ª ed. Barcelona, España: Marcial Pons, 2018.
- Polansky, Jonathan. Garantías constitucionales del procedimiento penal en entorno digital. 1.ª ed. Buenos Aires, Argentina: Hammurabi, 2020.
- Prometea. Inteligencia Artificial para transformar organizaciones públicas. 1.ª ed. Buenos Aires, Argentina: Astrea, 2019.
- Roth, Andrea. «Machine Testimony». Yale Law Journal 126, n.º 1 (enero de 2017): 1972-2053.
- Schwab, Klaus. La Cuarta Revolución Industrial. Barcelona, España: Editorial Debate, 2016.
- State of Wisconsin v. Eric L. Loomis, No. 2015AP157- CR (Supreme Court of Wisconsin 13 de julio de 2016).
- STC 62/2008, No. 3912-2005 (Tribunal Constitucional de España 26 de mayo de 2008).