INTELIGENCIA ARTIFICIAL Y ADMINISTRACIÓN DE JUSTICIA: LA POLICÍA PREDICTIVA Y LA JUSTICIA PREDICTIVA

Coloquio preparatorio del XXI Congreso Internacional de Derecho Penal de la Asociación Internacional de Derecho Penal

Directores: Javier A. De Luca y María Ángeles Ramos
Coordinadores: Francisco Figueroa, Hernán Kleiman y Manuela Parra

DOCTRINA

Poder Judicial de la Ciudad de Buenos Aires
Consejo de la Magistratura

JUS BAI RES EDITORIAL

Inteligencia Artificial y administración de Justicia: la policía predictiva y la justicia predictiva

Association Internationale de Droit Pénal
International Association of Penal Law
Asociación Internacional de Derecho Penal

www.editorial.jusbaires.gob.ar
editorial@jusbaires.gob.ar
fb: /editorialjusbaires
Av. Julio A. Roca 534 [C1067ABN]
+5411 4011-1320

Sello Buen Diseño argentino

Inteligencia artificial y administración de justicia : la policía predictiva y la justicia predictiva / Javier Augusto De Luca ... [*et al.*]. - 1ª ed. - Ciudad Autónoma de Buenos Aires: Editorial Jusbaires, 2025.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-768-421-6

1. Derecho de la Informática. I. De Luca, Javier Augusto
CDD 342

Hecho el depósito previsto según Ley Nº 11723

Declarada de interés por la Legislatura de la Ciudad Autónoma de Buenos Aires.
Res. Nro. 543-2018

Poder Judicial de la Ciudad de Buenos Aires
Consejo de la Magistratura

Autoridades

Presidenta

Karina Leguizamón

Vicepresidente 1º

Horacio Corti

Vicepresidente 2º

Manuel Izura

Consejeros

Lorena Clienti
Martín Converset
Luis Duacastella Arbizu
Marcelo Meis
Jorge Rizzo
Gabriela Zangaro

Secretaria de Administración General y Presupuesto

Genoveva Ferrero

ÍNDICE

WORKSHOP: INTELIGENCIA ARTIFICIAL Y LAS REGLAS DEL DERECHO

Prólogo e introducción

Javier Augusto De Luca*

Durante los días 28 al 31 de marzo de 2023 se desarrolló, en el Salón Azul de la Facultad de Derecho de la Universidad de Buenos Aires, el Coloquio de la Asociación Internacional de Derecho Penal (AIDP) bajo el lema "Inteligencia Artificial y Administración de Justicia: la policía y la justicia predictivas".

Los coloquios son seminarios o congresos más pequeños que se celebran en el lapso que transcurre entre congresos generales de la AIDP, los cuales se llevan a cabo cada cinco años. Al finalizar cada congreso general se decide cuál será el tema de derecho penal a tratar en el próximo, el cual será preparado y discutido en cuatro coloquios a desarrollarse en distintos países. En cada uno de esos coloquios se analiza el tema elegido bajo una de las cuatro perspectivas: derecho penal parte general, derecho penal parte especial, derecho procesal penal y derecho internacional penal.

En el último congreso general en Roma de 2019, se decidió que el tema del próximo congreso en París de 2024, sería el de Inteligencia Artificial y Derecho Penal. El grupo argentino de la AIDP fue elegido para organizar uno de los coloquios en la Argentina referido a los aspectos procesales penales, del cual da cuenta este libro.

Al momento de celebrarse el nuestro, ya se contaba con las conclusiones del coloquio sobre derecho penal parte general celebrado en Siracusa, Sicilia, Italia, a fines de 2022. Y al momento que escribo esto restan realizarse los coloquios de parte especial y derecho internacional, pospuestos por los dos años de la pandemia de COVID-19 que se apropió del mundo.

Los coloquios son dirigidos por un/a relator/a designado por la AIDP que tiene la tarea de redactar un proyecto de introducción y

* Doctor en Derecho. UBA. Profesor titular asociado de Derecho Penal y Procesal Penal, Facultad de Derecho, UBA. Presidente del Grupo Argentino de la Asociación Internacional. Fiscal General ante la Cámara Federal de Casación Penal.

consideraciones sobre el tema y proponer conclusiones o recomendaciones. En nuestro caso, este proyecto fue redactado por la profesora francesa Juliette Lelieur y fue presentado al público que asistió al encuentro, donde se discutieron todas las propuestas, palabra por palabra, hasta llegar a un consenso general sobre la redacción. Durante los tres días, participaron y asistieron más de cuarenta profesores y juristas extranjeros, y más de cien argentinos de distintas partes del país. Los puntos de discusión se votaron entre los asistentes de manera democrática. Al final se arribó a las conclusiones que se publicarán en esta obra. Esas recomendaciones serán discutidas, junto con las de los otros tres coloquios, en el congreso general a celebrarse en 2024, y aprobadas en asamblea general de la AIDP.

Aunque no existe una definición única y vinculante para el mundo jurídico, la inteligencia artificial puede caracterizarse como un conjunto de teorías y técnicas utilizadas para crear máquinas capaces de simular la inteligencia humana. El proceso de "toma de decisiones" es complejo y generalmente se asemeja al de una "caja negra" dentro de la cual no se puede conocer cómo ocurren las cosas.

Si los algoritmos son capaces de aprender y realizar nuevas destrezas, es posible sostener jurídicamente que cortan el cordón umbilical de sus creadores y, con ello, que pasan a ser difícilmente previsibles y controlables. Luego, esas dificultades se podrían extender y concretar en las investigaciones y los procesos penales, con afectación del derecho de defensa y falta de certeza sobre las imputaciones a las personas que realmente corresponden.

Se están utilizando estos mecanismos para prevenir o detectar delitos. Se basan en la evaluación de riesgos, la llamada policía predictiva. Se dice que pueden detectar patrones de probable comportamiento anormal o atípico. Así ya viene ocurriendo en la detección de posibles fraudes mediante el análisis de transacciones, y en sus conexiones con dispositivos de video-vigilancia en lugares públicos.

Los reconocimientos biométricos, los entrecruzamientos de datos para la búsqueda de prófugos, los jueces robots, la prueba de relación de causalidad en la prueba informática, la preservación e intangibilidad de esa clase de prueba, su reproducción fiel en los juicios para asegurar el derecho de defensa, las "expediciones de pesca" en las búsquedas automatizadas de información, el uso de las pruebas así ob-

tenidas mediante la violación del secreto de las comunicaciones y los papeles privados, el uso de drones para captar imágenes que invaden la intimidad, los errores en la individualización de blancos, de enemigos y de posesión de armas, la consagración como delitos de determinadas conductas que no lo son en todos los países y la posibilidad de una persecución penal internacional que viola la soberanía de los últimos (como puede ocurrir con la pedofilia y el *grooming* por internet), las pruebas surgidas de la *Deep Web*, etc. también fueron objeto de discusión y debate.

Si bien no fue tema concreto del Coloquio, los asuntos de derecho penal sustantivo surgieron en todo momento, porque están imbricados con los de derecho procesal de un modo que es imposible separarlos.

Para colmo se ha comprobado que los sistemas de IA no son completamente fiables ni tampoco neutros. Según ya lo preveía nuestra relatora del Coloquio, los errores derivan de la mala calidad de los datos utilizados o a la forma en que están programados los algoritmos, o a la existencia de falsos positivos/negativos en las correlaciones. Seguimos dependiendo de cómo fueron aprendiendo los sistemas, porque estos en realidad son influidos por las debilidades humanas, como son los algoritmos xenófobos, racistas, misóginos, los que parten de un discurso hegemónico que enfocan sus respuestas desde un único lugar que se asume como "correcto", etc.

Estos y otros temas de suma y actual importancia encontrarán los lectores en el libro que se me ha conferido el honor de prologar. Descuento que será de su agrado y de gran utilidad porque aquí se exponen las bases de lo que se viene, que está en constante desarrollo y a lo que los juristas no podemos ser indiferentes.

Buenos Aires, octubre de 2023

Participaciones locales

Los sistemas informáticos asistidos por Inteligencia Artificial en la administración de justicia penal argentina*

Carlos Christian Sueiro**

Introducción

El presente trabajo aborda la introducción de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal de la República Argentina.

El advenimiento de herramientas asistidas por inteligencia artificial (IA) se produce tanto en las áreas de prevención de delitos, dando lugar a la denominada "policía predictiva"; como en el servicio de administración de justicia propiamente dicho, a través de asistentes informáticos inteligentes destinados auxiliar a abogados particulares, defensores, querellantes, fiscales y jueces, a través de la elaboración automatizada de proyectos de escritos, dictámenes, recursos, y permitiendo efectuar incluso un análisis computacional del derecho en relación a las sentencias dictadas por los diversos tribunales orales criminal, dando inicio así al campo de la llamada "justicia predictiva".

El trabajo se divide en tres apartados o ejes temáticos centrales de análisis, en torno al estudio de la implementación de sistemas

---

* Este artículo es la versión extendida de la exposición realizada en la mesa redonda *"Inteligencia artificial en la actualidad Argentina"* del Coloquio Internacional de Derecho Penal *"Inteligencia Artificial y Justicia Penal"*, organizado por la Asociación Internacional de Derecho Penal (AIDP), llevado a cabo del 28 al 31 de marzo de 2023, en la Facultad de Derecho de la Universidad de Buenos Aires (UBA).

** Abogado, Diploma de Honor (2001), Especialista en Derecho Penal (2013), y Doctor en Derecho Penal (2021) por la Facultad de Derecho de la Universidad de Buenos Aires (UBA). Se desempeña como Profesor Adjunto en tema *"Criminalidad Informática y Prueba Digital"*. Profesor Adjunto (Int.) de multiples universidades. Coordinador del Centro de Estudios en Ciberseguridad y Protección de Datos (CECIB) de la Universidad CEMA (UCEMA). Se desempeña profesionalmente como Secretario Letrado de la Defensoría General Adjunta de la Nación.

informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal de la República Argentina.

En el primer apartado nos adentraremos en el estudio de "la inteligencia artificial y su impacto en la administración de justicia penal de la República Argentina", analizando la implementación de sistemas informáticos asistidos por inteligencia artificial (IA), tales como Sherlock Legal, Prometea y el Sistema de Reconocimiento Facial de Prófugos (SRFP) de la Ciudad Autónoma de Buenos Aires (CABA) y su proyección en otras provincias como Buenos Aires, Córdoba, Mendoza, Salta y Santa Fe.

En el segundo acápite del trabajo se abordará cual es el "marco legal en la República Argentina para la implementación de sistemas informáticos asistidos por inteligencia artificial (IA)".

En esta área analizaremos qué marco normativo se encuentra vigente en la República Argentina para la introducción de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal.

Se relevarán en este apartado los códigos procesales penales nacionales, federales y provinciales, como así también las leyes especiales, a fin de conocer, qué marco regulatorio a nivel legal existe en torno a los sistemas informáticos asistidos por inteligencia artificial (IA).

Finalmente, en el tercer y último apartado del trabajo, abordaremos cuáles son los "requisitos para la validación legal del empleo de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia".

Por último, de las conclusiones parciales y provisorias de cada apartado, daremos lugar a las conclusiones generales del trabajo.

La inteligencia artificial y su impacto en la administración de justicia penal de la República Argentina

La inteligencia artificial abre en la historia de la humanidad un sendero inexplorado y radicalmente excepcional, que sin lugar a dudas constituirá un punto de inflexión en nuestra evolución como especie.

Como bien expresa el filósofo y tecnólogo de origen francés, Éric Sadin, en nuestro presente nos encontramos frente a lo que él denomina "la emergencia de un nuevo régimen de verdad",[1] en el cual la inteligencia artificial juega un papel prioritario y estructural en la construcción de nuestras nuevas verdades relativas como especie. No será ya la inteligencia humana la destinada a evaluar caudales astronómicos de información exponencialmente crecientes en todas las áreas del desarrollo humano, sino la inteligencia artificial (IA).

En palabras del propio autor,

> Los sistemas de inteligencia artificial están llamados a evaluar una multitud de situaciones de todo orden, las necesidades de las personas, sus deseos, sus estados de salud, los modos de organización en común, así como una infinidad de fenómenos de lo real.[2]
>
> La posición del autor galo, también es compartida por la del filósofo sur coreano radicado en Alemania, Byung-Chul Han, para quien
>
> hoy se está produciendo de forma silenciosa un nuevo cambio de paradigma. El giro antropológico copernicano, que había elevado al hombre a productor autónomo del saber, es reemplazado por un giro dataísta. El hombre debe regirse por datos. Abdica como productor de saber y entrega su soberanía a los datos. El dataísmo pone fin al idealismo y al humanismo de la Ilustración. El hombre ha dejado de ser sujeto cognoscente soberano, autor del saber. Ahora el saber es producido maquinalmente. La producción de saber impulsada por datos se hace sin sujeto humano ni conciencia. Enormes cantidades de datos desbancan al hombre de su puesto central como productor de saber. Él mismo se ha atrofiado reduciéndose a un conjunto de datos, a una magnitud calculable y manejable.[3]

Por ello, resulta lógico que los sistemas informáticos asistidos por inteligencia artificial (IA), comiencen a implementarse también en el área de la administración de justicia

Como es lógico, el derecho y el sistema de administración de justicia no han prescindido de ella, y sistemas informáticos asistidos por

1. Sadin, Eric, *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical,* Buenos Aires, Caja Negra, 2020.

2. Ibídem, p. 95.

3. Han, Byung-Chul, *La desaparición de los rituales. Una topología del presente*, Barcelona, Herder, 2020, p. 105.

algoritmos de inteligencia artificial (IA), han irrumpido en la administración de justicia penal de varios Estado nacionales.

Numerosos estados nacionales han implementado desde la década pasada sistemas informáticos asistidos por algoritmos de inteligencia artificial (IA), tanto en la administración de justicia penal bajo la modalidad de "justicia predictiva", como en las áreas de prevención del delito, en el campo de la denominada "policía predictiva".

Entre aquellos estados nacionales que ha incorporado sistemas informáticos asistidos por inteligencia artificial (IA) en las áreas de justicia predictiva y policía predictiva, pueden mencionarse a modo de ejemplo: la República Popular de China,[4] Alemania,[5] España,[6] Estados

4. La República Popular de China posee un sistema omnipresente de vigilancia electrónica asistida por inteligencia artificial (IA) denominado *"Sense Time"* o *"Programa de crédito social ciudadano"* que es un sistema de policía predictiva, que abarca todas las áreas de la vida las personas. Cf. Sueiro, Carlos Christian, *Vigilancia electrónica y otros modernos medios de prueba*, 2ª Edición, Buenos Aires, Editorial Hammurabi, 2019, pp. 229-232.

5. Alemania cuenta con múltiples sistemas informáticos asistidos por inteligencia artificial, tanto en su función de Policia Predictiva como en el área de Justicia Predictiva. Sistemas informáticos asistidos por inteligencia artificial en el área de policía predictiva: 1.-PreCobs; 2.-KLB Operativ; 3.- KrimPro; 4.- Pre MAP; 5.- SKALA; 6.- RADAR-iTE; 7.- PNR (Passanger Name Record); 8.- Hessen Data; 9.- X-SONAR; 10.- ERAME. Sistemas informáticos asistidos por inteligencia artificial en el área de justicia predictiva: 1.- Herramientas informáticas asistidas por IA sobre evaluación de reincidencia; 2.- Base de datos para la elaboración automatizada de proyectos de sentencia. Sprenger, Johanna; Brodowski, Dominik, *Artificial Intelligence in the Administration of Criminal Justicie in Germany*, Asociación Internacional de Derecho Penal, 2023. Véase para una primera aproximación al tema: Molinas, Juan, *Procesos penales predictivos. La influencia de la inteligencia artificial y sus posibles límites,* Buenos Aires, Editorial Hammurabi, 2021, p. 310-335.

6. España cuenta como sistema informático asistido por inteligencia artificial (IA) en el área de la policía predictiva a VERIPOL. El sistema escanea la denuncia tomada en la comisaria o seccional policial y compara esa denuncia con su base de datos. Allí distingue los rasgos que pueden hacer que una denuncia sea validada como verdadera o clasificada como falsa. Por ejemplo, toma en consideración para validarla como verdadera la existencia de testigos o pruebas *in situ* en el lugar del hecho. Para su desarrollo han trabajado junto a las fuerzas de seguridad de España, la Universidad Complutense de Madrid, la Universidad Carlos III de Madrid, la Universidad de Roma *La Sapienza* y el Ministerio del Interior del Gobierno de España. Webedia Brand Service, *Precrime. ¿Cómo se utiliza la IA para la detección de crímenes futuros?*, Xataka Huawei, 2018. Disponible en: https://iahuawei.xataka.com/precrimen-como-se-utiliza-ia-para-deteccion-crimenes-futuros [Fecha de consulta: 01/03/2024].

Unidos de América,[7] Países Bajos u Holanda,[8] Reino Unido de Gran Bretaña e Irlanda del Norte.[9]

La República Argentina también se sumó a la lista de países que incorporaron la inteligencia artificial en la administración de justicia penal y en las áreas de prevención del delito.

En la pasada década, más precisamente a partir del año 2016, comenzaron a irrumpir en el sistema de administración de justicia de la República Argentina, los primeros sistemas informáticos asistidos por algoritmos de inteligencia artificial (IA).

Los tres sistemas informáticos asistidos por algoritmos de inteligencia artificial (IA), operativos en la República Argentina son: 1.- Sherlock Legal; 2.- Prometea; 3.- Sistema de Reconocimiento Facial de Prófugos (SRFP).

---

7. Los Estados Unidos de América (EE.UU) al igual que Alemania, posee múltiples sistemas informáticos asistidos por inteligencia artificial (IA), tanto en su función de Policía Predictiva como en el área de Justicia Predictiva. Sistemas informáticos asistidos por inteligencia artificial en el área de policía predictiva: 1.- CompSat; 2.- PredPol; 3.- Data Mind. Sistemas informáticos asistidos por inteligencia artificial en el área de justicia predictiva: 1.- IBM Wtason Debater; 2.- COMPAS -*Correctional Offender Management Profiling for Alternative Sanctions*-. Al respecto se puede consultar: Nieva Fenoll, Jordi, *Inteligencia artificial y proceso judicial. Proceso y Derecho*, Buenos Aires, Editorial Marcial Pons, 2018, p. 30.

8. Los Países Bajos u Holanda poseen a la fecha tres sistemas informáticos asistidos por inteligencia artificial (IA), en el área de la policía predictiva: 1.- Cold Case; 2.- Sweetie; y 3.- VALCRI (*Visual Analytics for sense making in Criminal Intelligence analysis*). Cerezoli Carlos, Alberto, *Estudio del agente encubierto informático como especial técnica de investigación del ordenamiento jurídico de España,* Buenos Aires, Editorial Hammurabi, p. 130. También consultar, Gasparini Neves, Estefania., *El agente encubierto informático en la República Argentina,* Buenos Aires, Editorial Hammurabi, 2020, pp. 163-173. Estévez Mendoza, Lucana, *Los derechos fundamentales ante las nuevas tecnologías: protección o vulneración a la luz de HART y VALCRI,* Buenos Aires, Editorial Thomson Reuters, 2019.

9. Reino Unido de Gran Bretaña e Irlanda del Norte, posee varios sistemas informáticos asistidos por inteligencia artificial en el área de policía predictiva y de la justicia predictiva. Sistemas informáticos asistidos por inteligencia artificial en el área de policía predictiva: 1.- NDAS (*National Data Analytics Solution*); 2.- HART (*Harm Assessment Risk Tool*). Sistemas informáticos asistidos por inteligencia artificial en el área de justicia predictiva: 1.- *Questmap*. Sobre los sistemas informáticos asistidos por IA en el Reino Unido de Gran Bretaña e Irlanda del Norte se sugiere consultar, las obras de Nieva Fenoll, Jordi, *Inteligencia artificial y proceso judicial. Proceso y Derecho*, Buenos Aires, Editorial Marcial Pons, 2018, pp. 26-30; y en particular sobre el sistema HART, confrontar Estévez Mendoza, Lucana, "Los derechos fundamentales ante las nuevas tecnologías: protección o vulneración a la luz de HART y VALCRI*", Revista de Derecho Penal y Criminología*, Nº 10, Buenos Aires, Editorial Thomson Reuters, 2019, p. 63.

Los dos primeros de los sistemas informáticos asistidos por algoritmos de inteligencia artificial (IA), Sherlock Legal (2016) y Prometea (2017), se encuentran orientados a la asistencia mediante inteligencia artificial (IA) en el campo de la administración de justicia, encontrándose insertos dentro del área de la justicia predictiva.

El primero de ellos, Sherlock Legal, fue diseñado, desarrollado y puesto en funcionamiento por el sector privado, más precisamente por Albremática S.A. dueña de la editorial digital, ElDial.com.

El segundo sistema, Prometea, por el contrario, es de origen estatal y fue elaborado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires (CABA).

Por su parte, el tercer sistema informáticos asistidos por algoritmos de inteligencia artificial (IA), el Sistema de Reconocimiento Facial de Prófugos (SRFP), se encuentra inmerso en el campo de la denominada Policía Predictiva, habiendo sido diseñado, elaborado, desarrollado y puesto en funcionamiento por el Gobierno de la Ciudad Autónoma de Buenos Aires.

Este último sistema informático asistido por inteligencia artificial (IA) en el área de la denominada Policía Predictiva, como lo es el Sistema de Reconocimiento Facial de Prófugos (SRFP), en los últimos cuatro años se proyectó por fuera de la Ciudad Autónoma de Buenos Aires; y este modelo ha dado origen a otros sistemas de reconocimiento facial, en las provincias tales como Buenos Aires, Córdoba, Mendoza, Salta y Santa Fe.

A continuación, referiremos a cada uno de estos tres sistemas informáticos asistidos por inteligencia artificial (IA) los cuales están siendo utilizados en la administración de justicia de la República Argentina, tanto en el campo de la denominada justicia predictiva, como en el área de la policía predictiva.

Sherlock Legal

A comienzos del año 2016, por iniciativa del sector privado argentino, más específicamente de la Editorial ElDial.com y bajo la dirección del Profesor Emérito en Derecho de Alta Tecnología de la Universidad Católica Argentina (UCA), Dr. Horacio R. Granero, se creó el programa asistido por inteligencia artificial denominado Sherlock Legal.

El asistente jurídico auxiliado por inteligencia artificial denominado Sherlock Legal, está creado sobre la base de datos jurisprudencial que posee la editorial, y emplea un *software* de procesamiento de lenguaje natural (NLP) asistido por IA, que le permite la extracción automatiza de conocimiento de textos jurídicos.

El programa de procesamiento de lenguaje natural (NLP) asistido por IA, que emplea Sherlock Legal, es un derivado del *Watson Legal de IBM*,[10] debido a que, para el desarrollo de este asistente jurídico inteligente, Albrematica S.A., dueña de la Editorial ElDial.com, realizó una alianza estratégica con IBM para que su *Biblioteca Jurídica Online de ElDial.com* pudiera ser accedida y analizada por el *software* asistido por IA.[11]

El asistente jurídico inteligente Sherlock Legal, gracias al *software* de procesamiento de lenguaje natural (NLP) asistido por IA, puede lograr entre sus funciones: 1.- categorización de textos jurídicos; 2.- agrupamiento de textos de acuerdo a su contenido; 3.- extracción de información dentro del texto, distinguiendo datos valiosos dentro del texto no estructurado; 4.- identificación, mediante la extracción de nombres de personas físicas y jurídicas; 5.- identificación de relaciones y vínculos entre los sujetos identificados; 6.- análisis emocional y sentimental a través del empleo de gramática, semiótica, psicología.

Como bien expresa el director a cargo del proyecto, el Dr. Horacio Granero, Sherlock Legal fue

> ... creado para ser un asistente del abogado, a través de una interfaz gráfica dinámica, intuitiva y sencilla donde se efectúan las preguntas en "lenguaje natural" (por ejemplo "En el caso de una separación matrimonial los hijos deben quedar con la madre?") no con "descriptores" o "lógica boolena" mencionada anteriormente. Mediante algoritmos generados al efecto se analizan sintácticamente y se interpretan los precedentes

10. IBM Watson Legal es una herramienta de inteligencia artificial (IA), de origen estadounidense diseñada y desarrollada por IBM, a la que se le plantea un tema jurídico de controversia o debate, y procede al análisis de los textos legales disponibles en internet sobre la materia. En base a toda la información colectada en la búsqueda, selecciona los argumentos que se presentan como más sólidos para expresar una conclusión jurídica con un lenguaje accesible. Cf. Nieva Fenoll, Jordi, *Inteligencia artificial y proceso judicial. Proceso y Derecho*, Buenos Aires, Editorial Marcial Pons, 2018, pp. 30 y 115.

11. Cf. Granero, Horacio, *"La inteligencia artificial aplicada al Derecho y el dilema de los algoritmos inteligentes"*, Buenos Aires, Editorial Hammurabi, pp. 38-39.

> judiciales con el fin de encontrar los fragmentos relacionados con la pregunta formulada que el programa considere más relevantes. Posteriormente, Sherlock despliega un grupo de las distintas respuestas que considera pertinentes, generándose gráficos que indican los porcentajes de aceptación o rechazo con la pregunta efectuada, dando, finalmente, su opinión en forma automática sobre la probabilidad que ésta sea afirmativa o negativa con relación a la consulta efectuada. El sistema si bien en esta primera etapa fue diseñado para ser aplicable a jurisprudencia, el desarrollo puede ser utilizado con cualquier base de datos estructurada y aún no estructurada.[12]

Actualmente, el proyecto se encuentra en etapa de escalamiento a otras ramas del Derecho, y ha recibido una mención honorífica en el Desafío Cognitiva celebrado en Costa Rica en Diciembre de 2017.

En septiembre de 2020, el equipo de I+D de Albrematica comenzó analizar la posibilidad de mejora de respuesta y potenciación de las habilidades de Sherlock Legal.

Por ello, Albrematica S.A. dueña de la Editorial ElDial.com, realizó una segunda alianza estratégica, en esa oportunidad con Open AI, el desarrollador de ChatGPT, con el objetivo de comenzar el desarrollo de una segunda versión de Sherlock Legal.

Más precisamente,

> ... en noviembre de 2020 Albrematica obtuvo de la empresa Open AI, creadora de GPT-3 la autorización para efectuar pruebas en su plataforma, avocándose el equipo técnico a su análisis, arribando a resultados auspiciosos por las capacidades predictivas demostradas hasta el momento de la herramienta.[13]

En el evento de celebración por el veinticinco (25) aniversario de ElDial.com, realizado en el octubre de 2022, se exhibió como Sherlock Legal puede ser operado no solo mediante interfaz de teclado o escritura, sino mediante comando de voz, lo cual permite que el asistente

12. Granero, Horacio, "La inteligencia artificial entiende el lenguaje "talcahuanensis". Impacto de los lenguajes de información legal, artículo publicado procesamiento de lenguaje natural (NLP) en la recuperación", Buenos Aires, Editorial El Dial.com, 04/03/2020, p.4.

13. Granero, Horacio, *GPT-3 y el futuro de la abogacía*, Buenos Aires, Editorial Hammurabi, 2021, p. 201.

jurídico auxiliado por inteligencia artificial (IA) pueda ser operado por personas con algún tipo de discapacidad.

Asimismo, Sherlock Legal también puede ser operado a través de teléfonos celulares inteligentes (*smartphone*) o terminales móviles multiplataformas convergentes de TIC, pudiendo ser consultado a través de la aplicación de telefonía instantánea WhatsApp.

El asistente jurídico inteligente auxiliado por IA, Sherlock Legal, consultado sobre un caso legal o tema jurídico, por medio de la aplicación de telefonía instantánea WhatsApp, otorgará respuesta a la consulta operando como un *chatbot* y enviando la respuesta en forma casi instantánea a la casilla de WhatsApp.

El asistente jurídico inteligente auxiliado por IA, Sherlock Legal de origen nacional, es muy similar a su vez a los asistentes jurídicos inteligentes Kleos diseñado por la firma anglo-neerlandesa Wolters Kluwer, a Legal One de Editorial La Ley, bajo la corporación transnacional anglo-canadiense, Thomson Reuters, y a Lex Machina de la empresa estadounidense, Lexis Nexis.

A su vez, también puede ser comparada como herramienta asistida por inteligencia artificial (IA), destinada a facilitar la labor de los abogados liberales con sistemas de mejora de argumentación jurídica y elaboración de alegatos o conclusiones legales, como los sistemas Questmap[14] de origen británico, y, Ross Intelligence,[15] oriunda de Canadá.

Prometea

> Prometea es una IA creada en Argentina, en el ámbito del Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires. El sistema fue ideado e implementado pensado en la optimización del servicio de justicia,

14. Nieva Fenoll, Jordi, *Inteligencia artificial y proceso judicial. Proceso y Derecho*, Buenos Aires, Editorial Marcial Pons, 2018, pp. 29-30.

15. Ross Intelligence es otra herramienta de inteligencia artificial (IA) semejante a *IBM's Watson Debater*, desarrollada en Canadá. Opera seleccionando legislación, jurisprudencia y doctrina, sobre un tema jurídico especifico formulado en forma de caso o hipótesis de hecho. Luego elabora una respuesta jurídica, en base al material seleccionado, realiza una argumentación jurídica, y calcula la tasa de probabilidad de éxito del planteo frente diversos estrados judiciales. Cf. Nieva Fenoll, Jordi, *Inteligencia artificial y proceso judicial. Proceso y Derecho*, Buenos Aires, Editorial Marcial Pons, 2018, p. 115.

con el fin de agilizar exponencialmente los procesos judiciales en beneficio del ciudadano.[16]

Desde comienzos del año 2017, el Ministerio Público Fiscal de la Ciudad de Buenos Aires, comenzó a explorar las nuevas tecnologías aplicadas al Derecho y al sistema de administración de Justicia, en cumplimiento con los lineamientos de la Planificación Estratégica 2017-2019.[17]

En virtud de ello, más precisamente a inicios del mes agosto de 2017, se comenzó con el desarrollo del sistema de inteligencia artificial Prometea.

Finalmente, el sistema de inteligencia artificial Prometea fue presentado en sociedad el miércoles 22 de noviembre de 2017, en el marco del "Congreso Internacional de Gobernanza inteligente e innovación inclusiva", en el Aula Magna de la Facultad de Derecho de la Universidad de Buenos Aires (UBA).[18] Técnicamente este sistema de inteligencia artificial es capaz de hacer algo imposible para cualquier ser humano, como operador del sistema de administración de justicia, que es leer, predecir, escribir y resolver un expediente judicial en 20 segundos, en promedio, y con una taza de acierto del 96%.[19] También tiene la capacidad de traducir el dictamen, escrito o recurso a tres idiomas, inglés, francés y portugués.

El sistema de inteligencia artificial (IA) Prometea se caracteriza por los siguientes cinco (5) aspectos distintivos:

1. Prometea es un sistema de inteligencia artificial que opera bajo supervisión humana.

   El sistema de inteligencia artificial, luego de cargados los hechos concretos y específicos del expediente, consulta su base de jurisprudencia, y mediante sistema de gramática, semióti-

16. Corvalán, Juan Gustavo, *Estados eficientes. La productividad del sector público*, Buenos Aires, Editorial Planeta, 2018, p. 260.

17. Corvalán, Juan Gustavo; Ciraudo, Denise, *Inteligencia artificial al servicio de la justicia penal, Contravencional y de faltas, "Prometea" en el ámbito de la Ciudad de Buenos Aires*, Buenos Aires, Montevideo, Editorial BdeF, 2018, p. 274.

18. Congreso Internacional de "Gobernanza Inteligente e innovación inclusiva. Desafíos y oportunidades para promover la efectividad de los Derechos en la cuarta revolución industrial", 22, 23 y 24 de Noviembre de 2017, Facultad de Derecho de la Universidad de Buenos Aires.

19. Corvalán, Juan Gustavo, "Estados eficientes. La productividad del sector público", en *Integración y Comercio*, 2018, Nº 44 (Julio), Buenos Aires, p. 261.

ca, derecho, elabora en 20 segundos promedio un escrito, dictamen, recurso, el cual pasa a ser controlado por un operador humano del sistema de administración de justicia, y luego de su revisión a la firma.

2. El sistema de IA Prometea emplea un modelo de pantalla integrada en el cual no hace falta cambiar de ventana para buscar información,[20] con lo cual la interfaz resulta más sencilla, amigable y eficiente.
3. Prometea también puede ser operado por personas que temporal o en forma permanente posean algún tipo de discapacidad para tipear la información. Gracias a el empleo de la denominada "inteligencia en la interfaz",[21] el usuario puede interactuar simplemente hablando a través de su sistema de reconocimiento de voz.
4. El sistema de inteligencia artificial Prometea permite "mejorar el nivel de previsibilidad, seguridad jurídica e igual. Por reducir la tasa o margen de error judicial, al otorgar a las mismas o similares circunstancias fácticas las mismas respuestas estatales".[22]
5. Prometea, bajo su sistema de automatización de tareas repetitivas y mecánicas, humaniza la labor de los miembros de la administración de justicia. Por ello, su empleo permite liberar tiempo valioso para que los integrantes de la administración de justicia pueden dedicarse a tareas más enriquecedoras y compleja como toma de decisiones.

Debe destacarse que el desarrollo de Prometea no está concluido sino que a este sistema se sumarán más aplicaciones con IA, un claro ejemplo es "la calculadora de prescripción", la cual es producto del desarrollos de las investigaciones realizadas en el instituto de Neurociencias y Derecho de la Fundación INECO (www.fundacionineco.org/institutos/inede) y en la Facultad de Derecho de la Universidad de Buenos Aires en el marco del Proyecto UBACyT sobre Neurociencia y Derecho.[23]

20. Ibídem, pp. 262-263.

21. Ídem.

22. Ibídem, p. 263.

23. Haissiner, Martin; Pastor, Daniel, *Neurociencias, tecnologías disruptivas y tribunales digitales*, Buenos Aires, Editorial Hammurabi, 2019, pp. 73-77.

El Sistema de Reconocimiento Facial de Prófugos (SRFP)

La República Argentina ya ha comenzado a operar sistemas informáticos asistidos por inteligencia artificial (IA) en el campo de la denominada policía predictiva desde el mes de abril del año 2019, con el empleo de un sistema de reconocimiento facial de prófugos (SRFP) asistido por un *software* de inteligencia artificial de origen ruso.

El sistema de vigilancia electrónica a través de reconocimiento facial asistido por inteligencia artificial (IA) no es una plataforma de convergencia de información tan completa y omnipresente como la utilizada por China con Sense Time, conocido como "Crédito social ciudadano"; en primer lugar no opera a nivel nacional, sino solo en la Ciudad Autónoma de Buenos Aires (CABA); y en segundo orden, no posee la convergencia aún de otros datos biométricos (reconocimiento de voz, huellas dactilares, iris ocular), o incluso mayor información biométrica proveniente de historias clínicas electrónicas.

El actual sistema de reconocimiento facial de prófugos (SRFP) asistido por algoritmos de inteligencia artificial forma parte del Sistema Público Integral de Videovigilancia y permite identificar los rostros de los acusado y condenados rebeldes o prófugos en menos de medio segundo.

Esto es posible gracias a una base de datos otorgada por el Co.Na.R.C. (Consulta Nacional de Rebeldías y Captura) que dispone de imágenes de los delincuentes y depende del Registro Nacional de Reincidencia, bajo la órbita del Ministerio de Justicia de la Nación.[24]

El sistema de reconocimiento facial de prófugos (SRFP) asistido por inteligencia artificial procesa las imágenes en línea (*online*) en tiempo real captadas por 300 de las 6963 cámaras instaladas por el gobierno porteño en las calles y estaciones de subterráneo.[25]

Por este motivo, las cámaras habilitadas con este *Software* serán rotadas según las necesidades y estrategias de investigación del Ministerio de Seguridad.

24. Gallo, Daniel, "Ya detuvieron a siete prófugos con el sistema de reconocimiento facial porteño", *La Nación*, 25/04/2019.

25. Ídem.

Las pruebas realizadas sobre el Sistema de reconocimiento facial por medio de inteligencia artificial arrojaron un rango de acierto superior al 93%.

Incluso el *software* puede mantener su elevado nivel de precisión aunque la persona intente modificar su apariencia visual por medio del uso de barba, bigote, cambios de corte de pelo, empleo de accesorios como anteojos, sombreros, gorras, capuchas, etcétera.26

El sistema de reconocimiento facial de prófugos (SRFP) a través de inteligencia artificial fue puesto en servicio operativo el miércoles, 24 de abril de 2019, y tal es su nivel de precisión que tan sólo a las 24 horas de su funcionamiento, el jueves 25 de abril, la Policía de la Ciudad "logró identificar a 11 prófugos de la Justicia y detener a siete de ellos".[27]

En la primera jornada de debut del sistema de reconocimiento facial de prófugos (SRFP) a través de inteligencia artificial, la vigilancia electrónica estuvo centrada en las estaciones de subte y traza vehicular del Metrobus de la Avenida 9 de Julio. El sistema tiene la capacidad de reconocer 750 rostros en simultáneo, lo cual lo hace sumamente idóneo y eficiente para detectar prófugos y rebeldes en lugares con alto nivel de concentración de personas como estaciones de subtes trenes, paradas de autobuses, terminales portuarias o aeropuertos. El sistema de reconocimiento facial de prófugos (SRFP), una vez identificado el sospechoso o rebelde, emite un alerta a los funcionarios encargados de las cámaras de vigilancia y a su vez también envía la información a los teléfonos celulares inteligentes (*Smartphone*) de los agentes policiales geolocalizados en mayor proximidad al sujeto identificado para coordinar el arresto.

A fin de evitar la vigilancia electrónica masiva de ciudadanos de la Ciudad Autónoma de Buenos Aires (CABA), el sistema de reconocimiento facial de prófugos (SRFP) por medio de inteligencia artificial solamente identifica o reconoce los rostros de personas incorporadas a la base de datos del Co.Na.R.C. (Consulta Nacional de Rebeldías y Captura).

Los funcionarios públicos encargados del sistema de vigilancia electrónica mediante reconocimiento facial de prófugos (SRFP) asistido por IA poseen el deber de confidencialidad, y la inserción de datos o imágenes, revelación de datos o acceso ilegitimo a este sistema

26. Ídem.
27. Ídem.

será una conducta típica de violación a una base de datos personales (Art. 157 bis CPN).

El empleo del sistema de vigilancia electrónica mediante reconocimiento facial asistido por IA –para el seguimiento de personas no incluidas en la base de datos del Co.Na.R.C. (Consulta Nacional de Rebeldías y Captura)– será considerado como una acción que puede encuadrar en alguno de los Delitos contra la Privacidad y Violación de Secretos, constituyendo ese obrar un claro accionar de inteligencia ilegal sobre el ciudadano.

El sistema de vigilancia electrónica mediante reconocimiento facial de prófugos asistido por inteligencia artificial se complementa con: 1.- El Sistema de reconocimiento electrónico de patentes, 2.- El Sistema Público Integral de Videovigilancia, y, 3.- La aplicación Mi Argentina App.[28]

El sistema de reconocimiento facial de prófugos (SRFP) fue mejorado durante el 2020, a raíz de la Pandemia de Coronavirus COVID-19, pues a las cámaras con reconocimiento facial se le ha sumado la capacidad de detección infrarroja de temperatura.[29]

Las cámaras térmicas dotadas de un *software* con inteligencia artificial (IA) permiten medir la fiebre a 20 personas a la vez, sin contacto físico, y con un margen de error menor a 0,3° grado celsius.

Las cámaras térmicas con IA son producidas por la empresa china Huawei. Para medir la temperatura de una persona sin contacto físico, las cámaras cuentan con un lente óptico, que identifica figuras humanas, y otro lente térmico que detecta la temperatura y la hace visible para el ojo humano.[30] También están instaladas en el aeropuerto inter-

28. La aplicación Mi Argentina viene a renovar la portabilidad digital de documentación esencial para los ciudadanos, como los son: 1.- el DNI, 2.- el Pasaporte, 3.- la Constancia de CUIL, 4.- el Certificado único de discapacidad, 5.- la credencial de transplantado y 6.- licencias de conducir.
Por medio de la aplicación podrán ser visualizados en los controles policiales todos estos documentos, incluso sin la necesidad de poseer conectividad. Se sugiere ver el sitio oficial del Estado Nacional www.argentina.gob.ar/aplicaciones-moviles.

29. Sueiro, Carlos Christian, *Vigilancia electrónica asistida por inteligencia artificial (IA)*, Buenos Aires, Editorial Hammurabi, 2020.

30. "Coronavirus: Como funcionan las cámaras con inteligencia artificial que se instalaron en Ezeiza", *Radio Mitre*, 13/03/2020. Disponible en: https://radiomitre.cienradios.com/coronavirus-como-funcionan-las-camaras-con-inteligencia-artificial-que-instalaron-en-ezeiza/ [Fecha de consulta: 01/03/2024].

nacional de Ezeiza.[31] Se suman otras cámaras térmicas colocadas en las estaciones de trenes de las estaciones Constitución, Retiro y Once, las cuales se complementan con otras TIC tales como: 1.- La aplicación COVID 19 (COVID App); 2.- El sistema de reconocimiento facial mediante inteligencia artificial instalado por el Gobierno de la Ciudad Autónoma de Buenos Aires en abril de 2019; 3.- Los arcos de detección de patentes de vehículos que permiten establecer un trazado digital de la trayectoria realizada por el automotor.

Asimismo, se agrega el sector privado que actúa en cooperación asimétrica con el Estado Nacional. Así pueden mencionarse: 1.- Las empresas prestadoras de servicios de Internet y telefonía móvil que, a requerimiento de la autoridad estatal, pueden brindar datos de geolocalización en tiempo real de la persona que porta el celular; 2.- Las empresas de recuperación de vehículos robados o rastreo de personas.[32]

En noviembre de 2020, por medio de la Ley Nº 5688 de la Ciudad Autónoma de Buenos Aires, al Sistema de Reconocimiento Facial de Prófugos se le adicionaron otros dos sistemas de policía predictiva, siendo ellos: 1.- El Sistema Preventivo y el Sistema Forense. 2.- El Sistema Integral de Seguridad Pública citadino.

Una de las problemáticas durante la Pandemia de Coronavirus COVID-19 radicó en el hecho de que el Sistema de Reconocimiento Facial de Prófugos (SRFP) fue empleado para reconocer a todos los ciudadanos que ingresaban o egresaban de la Ciudad Autónoma de Buenos Aires, a fin de hacer efectivo las políticas de Aislamiento Social, Preventivo y Obligatorio (ASPO) y Distanciamiento, Social, Preventivo y Obligatorio (DISPO).

Esto implicaba que el sistema de reconocimiento facial ya no solo accedía a la base de datos de CONARC (Consulta Nacional de Rebeldías y Captura), sino a la base de datos del Registro Nacional de las Personas (RENAPER).

31. TELAM, "Colocan cámaras infrarrojas de control de temperatura en Ezeiza", 12/03/2020. Disponible en: https://www.telam.com.ar/notas/202003/440307-coronavirus-ezeiza-camaras-infrarrojas-control-temperatura.html [Fecha de consulta: 01/03/2024].

32. *LoJack* Argentina está usando los rastreadores de los autos para denunciar que sus clientes no cumplen con la cuarentena. Para ello utilizan *Lo Jack* en los automóviles y Strix para el rastreo de personas.

Esta desnaturalización de su función original generó desde finales del año 2020 muchas controversias que duran hasta nuestros días.

Esta expansión de las funciones de reconocimiento facial masivo de la ciudadanía no suscitó un fuerte interés o debate en la opinión pública en general.

Sí debe destacarse que existen algunas organizaciones no gubernamentales (ONG) que resultan críticas respecto a falsos positivos, múltiples fallas y errores en la detección de personas sin antecedentes que presentaban semejanzas faciales con personas en situación de rebeldía, a las cuales se demoró como consecuencia del error del Sistema de Reconocimiento Facial de Prófugos (SRFP) de la Ciudad de Buenos Aires.

Con posturas críticas hacia el Sistema de Reconocimiento Facial de Prófugos (SRFP), podemos referir a la Asociación de Derechos Civiles (ADC), la cual –al implementarse el sistema– presentó una acción declarativa de inconstitucionalidad contra el Gobierno de la Ciudad de Buenos Aires bajo el argumento de que

> … el reconocimiento facial, cuando se aplica con fines de vigilancia policial, se convierte en una tecnología desproporcionada que, además de no contar con las bases legales apropiadas, afecta gravemente los derechos y las garantías constitucionales de todas las personas que desarrollan su vida en la ciudad. [33]

El sistema también fue fuertemente cuestionado por otras organizaciones no gubernamentales (ONG) como es el caso del Observatorio de Derecho Informático de Argentina (O.D.I.A.).[34]

El Observatorio de Derecho Informático Argentino (O.D.I.A.) presentó una acción de amparo contra el Gobierno de la Ciudad Autónoma de Buenos Aires en los términos del art. 14 de la Constitución CABA y las Leyes Nº 2145 y Nº 104, ante el Juzgado de 1° instancia en lo Contencioso Administrativo y Tributario N° 32, Secretaría N° 45, de la CABA.

---

33. Asociación por los Derechos Civiles, "El reconocimiento facial para vigilancia no pertenece a nuestro espacio público", 06/11/2019. Disponible en: https://adc.org.ar/2019/11/06/el-reconocimiento-facial-para-vigilancia-no-pertenece-a-nuestro-espacio-publico/ [Fecha de consulta: 01/03/2024].

34. "Por qué comprar un router Wi-Fi 6 y complementar al que las operadoras instalan de serie", en *Webedia Brand Services*, 18/12/2020. Disponible en: https://odia.legal/ [Fecha de consulta: 01/03/2024].

El juzgado hizo parcialmente lugar a la demanda y condenó al Gobierno de la Ciudad Autónoma de Buenos Aires (GCBA) a que complete la información ofrecida al Observatorio, en relación a la resolución 398/MJYSGC/2019, la cual aprobó la implementación del Sistema de Reconocimiento Facial de Prófugos (SRFP) en el ámbito de la Ciudad Autónoma de Buenos Aires (CABA).[35]

El Observatorio de Derecho Informático Argentino (O.D.I.A.) presentó una acción de amparo de acceso a la información pública que consistía en setenta y siete (77) preguntas, las cuales intentaban arrojar luz sobre el tipo de tecnología implementada y el funcionamiento en su conjunto del sistema de reconocimiento facial de prófugos (SRFP).

El universo de preguntas efectuadas, se dio respuesta parcialmente a ellas.

Así pueden distinguirse dos grupos de preguntas: 1.- Grupo de preguntas al que se le brindó una respuesta parcial, y 2.- grupo de consultas o preguntas al que no se le otorgó respuesta alguna.

Se otorgó una respuesta parcial a las siguientes consultas: 1.- consultas sobre cifrado de información y privacidad, 2.- alcance del reconocimiento facial, 3.- registro y archivo de información de menores de edad, 4.- alcance del convenio con la base de datos del RENAPER, 5.- funcionalidad de aparatos de repetición y alerta, 6.- método de detección de rostros de personas condenadas, 7.- Nivel de precisión en la detección de rostros, 8.- auditoría independiente del *Software*.

Mientras que, a criterio del magistrado a cargo, no se dio respuesta alguna a los siguientes interrogantes formulados: 1.- Cuáles son los protocolos de seguridad y confiabilidad de las capturas de imágenes faciales, 2.- Auditoría de borrado de datos, 3.- Identificación de personas no incorporadas a las bases de datos de CONARC y Registro Nacional de Reincidencia, 4.- Detección del porcentaje de falsos positivos, 5.- Determinación de agentes que reciben información confidencial.[36]

35. CCAyT CABA, "Observatorio de Derecho Informático Argentino O.D.I.A. c/GCBA s/ Acceso a la Información", Expte. Nº 182908/2020, 20/05/2020.

36. Vergara Vacarezza, Diego Alonso; Gamarra Calello, Santiago, *Sistema de reconocimiento facial en el proceso penal Anotación al fallo "Observatorio de Derecho Informático Argentino O.D.I.A c GCBS S/Acceso a la información",* Buenos Aires, Editorial Hammurabi, 2021, pp. 447 y 477.

Le decisión judicial consideró que la parte demandada dio respuesta a la mayoría de las preguntas formulada por la actora, tanto en sede administrativa como durante la tramitación de la causa.[37]

Posteriormente, el 12 de abril de 2022, el Juez a cargo del Juzgado de 1ra Instancia en lo Contencioso Administrativo y Tributario N° 2, Dr. Roberto Gallardo, a través del fallo Nº 783420/2022 "Observatorio de Derecho Informático Argentino O.D.I.A. Sobre otros procesos incidentales - Amparo – Otros",[38] dispuso la suspensión del sistema de reconocimiento facial de prófugos (SRFP).[39]

La decisión del titular a cargo del Juzgado de 1ra Instancia en lo Contencioso Administrativo y Tributario N° 2, Juez Roberto Gallardo, fue en respuesta a una acción de amparo de solicitud de información pública, solicitada en el año 2019 por el Observatorio de Derecho Informático Argentino (O.D.I.A.) y que recibió el apoyo posterior del Centro de Estudio Legales y Sociales (CELS).

Se afirma en la sentencia que el sistema de reconocimiento facial de prófugos (SRFP), no solo contaba con información de los prófugos por medio de la base de Consulta Nacional de Rebeldías y Captura (CONARC), sino que el Registro Nacional de las Personas (RENAPER) le había entregado al Ministerio de Seguridad y Justicia de la Ciudad Autónoma de Buenos Aires (MJySCABA) datos de otros ciudadanos que no poseían antecedentes penales y cuya finalidad legal no resulta clara.

Debe destacarse que el debate en la comunidad académica y científica, en torno al sistema de reconocimiento facial de prófugos (SRFP), se encuentra en una etapa inicial y comienzan a publicarse los primeros trabajos en torno al tema en libros y revistas jurídicas.[40]

---

37. CCAyT CABA, "Observatorio de Derecho Informático Argentino O.D.I.A. c/GCBA s/ Acceso a la Información", Expte. Nº 182908/2020, 20/05/2020.

38. CCAyT CABA "Observatorio de Derecho Informático Argentino O.D.I.A. Sobre otros procesos incidentales - Amparo – Otros", Expte. Nº 182908/2020, 20/05/2020.

39. "Otro fallo judicial polémico: suspenden el sistema de reconocimiento facial de la Ciudad Autónoma de Buenos Aires*", Diario Clarín*, 12/04/2022. Disponible en: Otro fallo judicial polémico: suspenden el sistema de reconocimiento facial en la Ciudad de Buenos Aires (clarin.com) [Fecha de consulta: 01/03/2024].

40. Riquert, Marcelo; Sueiro, Carlos Christian, *"Sistema Penal e Informática"*, Buenos Aires, Editorial Hammurabi, 2018. Como así también en la obra de Danesi, Cecilia, "Inteligencia artificial, tecnologías emergentes y derecho. Reflexiones interdisciplinarias", Buenos Aires, Editorial Hammurabi, 2020.

Lo cierto, es que el sistema de reconocimiento facial de prófugos (SRFP), pese a su momentánea suspensión judicial, es un modelo que tiende a replicarse en otras provincias de la República Argentina, como modelo de seguridad urbana orientado al campo de la policía predictiva.

Otros sistemas de reconocimiento facial asistidos por IA

Además del Sistema de Reconocimiento Facial de Prófugos (SRFP) de la Ciudad Autónoma de Buenos Aires (CABA), en la década pasada y en los primeros tres años de la década en curso, otras jurisdicciones provinciales han introducido sistemas de reconocimiento facial como mecanismo de policía predictiva.

La Asociación por los Derechos Civiles (ADC) ha relevado la incorporación de sistemas de reconocimiento facial en cinco (5) provincias, siendo ellas Buenos Aires, Córdoba, Mendoza, Salta y Santa Fe; y en al menos ocho (8) ciudades del país, además de la Ciudad Autónoma de Buenos Aires (CABA).

Veamos muy brevemente estos sistemas de reconocimiento facial como mecanismo de policía predictiva.

Provincia de Buenos Aires

En la Provincia de Buenos Aires al menos tres (3) ciudades han implementado sistemas de reconocimiento facial bajo la modalidad de policía predictiva.

- Municipio de Tigre

En el año 2019 comenzaron las tratativas para la implementación de un sistema de reconocimiento facial con fines de seguridad urbana, el cual se hizo operativo en el año 2020 durante la Pandemia de Coronavirus COVID-19.

El sistema utiliza un *software* desarrollado por la empresa tecnológica NEC, el cual mediante un algoritmo asistido por inteligencia artificial (IA) permite la identificación de personas con antecedentes penales, o bien, que se encuentren perdidas o extraviadas.

El sistema de reconocimiento facial con fines de seguridad urbana, incorporó diez (10) nuevas cámaras en puntos estratégicos del municipio con capacidad de reconocimiento facial en tiempo real.

El sistema de reconocimiento facial con fines de seguridad urbana es íntegramente monitoreado desde el Centro de Operaciones de Tigre (COT).[41]

- Partido de La Matanza

En el año 2020 el Poder Ejecutivo municipal, con apoyo del Ministerio de Seguridad de la provincia de Buenos Aires, comenzó con la implementación gradual y progresiva de un sistema de reconocimiento facial con fines de seguridad urbana, del cual no se han brindado detalles respecto del *software* que utiliza asistido por inteligencia artificial, ni se han especificado sus capacidades de reconocimiento y detección.[42]

- Ciudad de Mar del Plata

El Concejo Deliberante de la Ciudad de Mar del Plata aprobó en agosto de 2022 la iniciativa del Poder Ejecutivo municipal de incorporación de un *software* asistido por inteligencia artificial (IA).

Se trataría de un *software* que no solo permitiría la identificación facial, sino también la detección dinámica de conductas compatibles con la ejecución de hechos delictivos, incrementando así la capacidad predictiva de las fuerzas de seguridad de la ciudad marplatense.

Desgraciadamente, la información brindada por el Poder Ejecutivo municipal en relación al sistema de reconocimiento facial con fines de seguridad urbana es sumamente escasa.[43]

Córdoba

El gobierno de Córdoba implementó en octubre del año 2019 un Sistema de Reconocimiento Biométrico (SRB) en el marco del plan estratégico de seguridad.

El Sistema de Reconocimiento Biométrico (SRB) utiliza un *software* de reconocimiento facial asistido por inteligencia artificial que no ha sido explicitado o dado a conocer por las autoridades.

41. Sistema de Reconocimiento Facial del Municipio de Tigre, "*Con mi cara no*", Asociación por los Derechos Civiles (ADC). Disponible en: https://conmicarano.adc.org.ar/ [Fecha de consulta: 01/03/2024].

42. Ibídem, partido de La Matanza.

43. Ibídem, Ciudad de Mar del Plata.

El sistema emplea un grupo de cámaras fijas instaladas en puntos estratégicos y cámaras móviles instaladas en patrulleros de la policía de Córdoba.

El gobierno de la provincia de Córdoba se ha negado a brindar información y no dado respuesta alguna a dos pedidos de acceso a la información pública.[44]

Mendoza

En el año 2017 el gobierno de la provincia de Mendoza propuso en el marco de un plan estratégico de seguridad la implementación de un sistema de reconocimiento facial.

A mediados del año 2019 comenzó la incorporación de *software* asistido por inteligencia artificial (IA) en las cámaras de los móviles policiales de Mendoza.

Las autoridades provinciales se negaron a brindar información sobre el *hardware* y *software* utilizado por el sistema de reconocimiento facial, bajo el pretexto de que ello podría afectar la seguridad del sistema y explicitar vulnerabilidades.[45]

Salta

El gobierno de la provincia de Salta implementó, desde comienzos del año 2018 y hasta fines del año 2019, un sistema de reconocimiento facial que abarca toda la provincia.

Bajo el proyecto denominado "Salta inteligente", se instalaron mil cuatrocientas (1400) cámaras con capacidad de reconocimiento facial a lo largo de toda la provincia.

Las cámaras fueron instaladas en sitios estratégicos de la Ciudad de Salta, tales como la Zona del Cabildo, la peatonal Florida, y a lo largo de la provincia en Aeropuertos, terminales de ómnibus, estaciones de tren, proximidad de hospitales.[46]

44. Ibídem, Provincia de Córdoba.

45. Ibídem, Provincia de Mendoza.

46. Ibídem, Provincia de Salta.

Santa Fe

Durante la Pandemia de Coronavirus COVID-19, más precisamente en noviembre de 2020, el Ministerio de Seguridad de la Provincia de Santa Fe anunció la implementación de un Sistema de Reconocimiento Facial en la ciudad de Rosario, la cual en la actualidad se ve sumida en el caos de la narco-criminalidad, contando con la mayor tasa de homicidios al año en todo el país.

El gobierno nacional se comprometió a incrementar el gasto público en materia de seguridad, a fin de que en el año 2021 el sistema de reconocimiento facial estuviera operativo.[47]

Sin embargo, a la fecha, el sistema de reconocimiento facial de la Ciudad de Rosario no se ha concluido.

Marco legal en la República Argentina para la implementación de sistemas informáticos asistidos por inteligencia artificial (IA)

En relación a la regulación legal de la inteligencia artificial en la República Argentina, debe aclararse que en el sistema normativo argentino no existe una ley que defina inteligencia artificial, policía predictiva o justicia predictiva.

A la fecha no se encuentra regulada específicamente en los códigos procesales penales federales de la República Argentina, siendo ello una materia pendiente teniendo en consideración que, desde la incorporación al ordenamiento nacional del Convenio de Ciberdelincuencia de Budapest, el 22 de noviembre de 2017, por medio de la Ley Nº 27411, corresponde por mandato constitucional y convencional, adecuar el régimen procesal penal a los requerimientos de la sección segunda de dicho convenio.

En el análisis a nivel procesal penal de la inteligencia artificial (IA), resulta sumamente lógico y sensato, desde una perspectiva histórica comprender que el primer Código Procesal Penal de la Nación, no contemplara esta materia.

47. Ibídem, Ciudad de Rosario.

El primer Código de forma en materia penal de la República Argentina, el Código de Procedimiento en Materia Penal (Ley Nº 2372)[48] sancionado el 4 de octubre de 1888, y vigente hasta el año 1991, fue un Código Procesal Penal que implementaba un sistema de enjuiciamiento inquisitivo.

Durante la mayor parte de su período de vigencia las tecnologías de la información y la comunicación (TIC) ni siquiera se habían desarrollado, con lo cual resulta materialmente imposible la previsión de este tipo de prueba.

Respecto al segundo Código procesal en materia penal de nuestra Nación, el Código Procesal Penal de la Nación de la República Argentina (Ley Nº 23984),[49] vigente desde 1991 hasta la actualidad, tampoco contempló la introducción de ninguna definición de inteligencia artificial (IA), policía predictiva o justicia predictiva.

La razón de que al momento de su sanción, promulgación y puesta de entrada en vigencia, no se hayan incorporado al título de medios de prueba a la evidencia digital, obedece a que las TIC recién por la década de los 90 del pasado siglo XX estaban comenzando a ser conocidas. Considérese que recién a partir del año 1996, como consecuencia del arribo de Internet a nuestro país, comienzan a debatirse los primeros Proyectos de Ley sobre la incorporación de tipos penales al Código Penal de la Nación relacionados con la Criminalidad Informática.

De modo que, hasta el momento, nos hemos manejado con el criterio normativo de amplitud probatoria y la veracidad o seriedad de las pruebas de esa naturaleza o las adquiridas por medio de IA, han sido aceptadas y su valor probatorio dependerá del examen de peritos o expertos que las doten de autenticidad y seriedad. Es decir, se admite este tipo de pruebas como cualquier otra, de acuerdo a los principios de libertad probatoria y sana crítica racional.

El nuevo Código Procesal Penal Federal (Ley Nº 27063 modificado por Ley Nº 27482), no ha introducido ninguna regulación específica sobre nuevos medios de prueba puntuales destinados a la obtención de prueba digital alojada en diversos dispositivos electrónicos o equipos

48. Ley Nº 2372, Código de Procedimiento en Materia Penal, sancionada el 04/10/1988 y promulgada el 01/01/1989.

49. Ley Nº 23984, Código Procesal Penal de la Nación, sancionada el 21/08/1991 y promulgada el 04/09/1991.

informáticos, siendo ello una materia pendiente hasta nuestros días. Solo hace una referencia muy genérica a la incautación de datos en su artículo 151.[50]

Por consiguiente, menos aún se han incorporado normas destinadas a la regulación de la inteligencia artificial como medio de obtención de prueba en la administración de justicia.

La mayoría de las provincias de la República Argentina, teniendo en consideración de que cada una de ellas están facultadas a dictar su propio código procesal penal, hasta la fecha tampoco han adecuado su legislación procesal penal a los requerimientos del Convenio de Ciberdelincuencia de Budapest (2001); el cual, como ya se mencionó, se incorporó a nuestro ordenamiento jurídico por medio de la Ley Nº 27411 el 22 de noviembre de 2017.

Si bien el Convenio de Ciberdelincuencia de Budapest fue tomado como principio rector a nivel nacional para la reforma integral al Código Penal de la Nación en materia de criminalidad informática, realizada en el año 2008, a través de la Ley Nº 26388, no menos cierto resulta que hasta el momento no se ha tenido en consideración la sección segunda del convenio, para efectuar una reforma en materia procesal penal que permita la regulación de la prueba digital y de los medios de prueba específicos para su obtención, recolección, peritaje y conservación, ni a nivel nacional y tampoco masivamente por parte de las provincias.

50. Art. 151 CP "El juez podrá ordenar a requerimiento de parte y por auto fundado, el registro de un sistema informático o de una parte de éste, o de un medio de almacenamiento de datos informáticos o electrónicos, con el objeto de secuestrar los componentes del sistema, obtener copia o preservar datos o elementos de interés para la investigación, bajo las condiciones establecidas en el artículo 136.
Regirán las mismas limitaciones dispuestas para el secuestro de documentos.
El examen de los objetos, documentos o el resultado de la interceptación de comunicaciones, se hará bajo la responsabilidad de la parte que lo solicitó. Una vez secuestrados los componentes del sistema, u obtenida la copia de los datos, se aplicarán las reglas de apertura y examen de correspondencia.
Se dispondrá la devolución de los componentes que no tuvieran relación con el proceso y se procederá a la destrucción de las copias de los datos. El interesado podrá recurrir al juez para obtener la devolución de los componentes o la destrucción de los datos".

Tal es así que las provincias de Buenos Aires,[51] Catamarca,[52] Chaco,[53] Chubut,[54] Córdoba,[55] Entre Ríos,[56] Formosa,[57] Jujuy,[58] La Pampa,[59] La Rioja,[60] Mendoza,[61] Misiones,[62] Salta,[63] San Juan,[64] San Luis,[65] Santa Cruz,[66] Santa Fe,[67] Santiago del Estero[68] y Tierra del Fuego;[69] no han regulado nuevos medios de prueba propios de la era digital, para permitir la adquisición de evidencia electrónica por medio de inteligencia artificial (IA), policía predictiva o justicia predictiva.

Todos estos códigos procesales penales de las provincias previamente mencionadas suelen recurrir al principio de libertad probatoria, para permitir la aplicación analógica de un medio de prueba propio de la prueba física, para lograr la obtención de evidencia digital, lo cual debe ser atendido con cuidado, porque puede afectar las garantías constitucionales y convencionales de los imputados, tales como los derechos a la privacidad, intimidad, derecho a la autodeterminación

51. Ley Nº 11922, Código Procesal Penal de la Provincia de Buenos Aires.

52. Ley Nº 5097, Código Procesal Penal de la Provincia de Catamarca, sancionada en San Fernando del Valle de Catamarca el 03/07/2023.

53. Ley Nº 4538, Código Procesal Penal de Chaco, sancionada el 04/11/1998.

54. Ley Nº 5478, Código Procesal Penal de Chubut.

55. Ley Nº 8123, Código Procesal Penal de Córdoba, sancionada el 05/12/1991.

56. Ley Nº 9754, Procesal Penal de Entre Ríos, sancionada el 09/01/2007.

57. Ley Nº 696/87, Código Procesal Penal de la provincia de Formosa.

58. Ley Nº 5623, Código Procesal Penal de la provincia de Jujuy, sancionada el 05/11/2009.

59. Ley Nº 332, ordenada y concordada por la Ley Nº 713/95, Código Procesal Penal de la provincia de La Pampa.

60. Ley Nº 8774, Código Procesal Penal de la provincia de La Rioja.

61. Ley Nº 6730, Código Procesal Penal la provincia de Mendoza. No obstante la provincia de Mendoza hace tres años comenzó con un proyecto de reforma a los efectos de incorporar medios de prueba orientados a la obtención de evidencia electrónica.

62. Ley Nº 14, Código Procesal Penal de la provincia de Misiones, sancionada el 10/10/2013.

63. Ley Nº 7690, Código Procesal Penal de la provincia de Salta, 01/11/2011.

64. Ley Nº 754, Código Procesal Penal de la provincia de San Juan.

65. Ley Nº 5724, Código Procesal Penal de la provincia de San Luis.

66. Ley Nº 2424, Código Procesal Penal de la provincia de Santa Cruz.

67. Ley Nº 12734, Código Procesal Penal de la provincia de Santa Fe.

68. Ley Nº 6941, Código Procesal Penal de la provincia de Santiago del Estero.

69. Ley Nº 168, Código Procesal Penal de la provincia de Tierra del Fuego, sancionada el 19/08/1994.

informativa, derecho al anonimato, defensa en juicio del acusado, y prohibición de no autoincriminación forzada.

Solamente los códigos procesales penales de las provincias de Corrientes (Ley N° 6518),[70] Neuquén (Ley Nº 2784),[71] Río Negro (Ley Nº 5020)[72] y Tucumán (Ley Nº 8933),[73] han incorporado la prueba digital y nuevos medios de prueba propios y específicos de la era digital para su correcta obtención; pero pese a su actualización, no han introducido ninguna norma referente a la implementación de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia.

Sin embargo, pese a no existir hasta la fecha una ley específica que regule la implementación de inteligencia artificial (IA) en la administración de justicia en materia penal, todo sistema informático asistido por inteligencia artificial (IA) que se introduzca en la administración de justicia en calidad de policía predictiva o justicia predictiva, debe respetar las garantías prevista en la Constitución Nacional, las Convenciones y Tratados de Derechos Humanos, los principios plasmados en El Convenio de Ciberdelincuencia de Budapest (Ley Nº 27411) y con la Ley de Protección de Datos Personales (Ley Nº 25326).

No obstante, no basta con aclarar y reiterar que la inteligencia artificial (IA) aplicada en la administración de justicia penal –pese a no estar regulada aún legal y normativamente– debe ser respetuosa de la Constitución Nacional, las Convenciones y Tratados de Derechos Humanos,

70. Ley Nº 6518, Código Procesal Penal de Corrientes, sancionada 07/11/2019. Este código reguló los siguientes medios de prueba para la adquisición de prueba digital: 1.- Vigilancia de las comunicaciones (Art. 216); 2.- Vigilancia remota sobre equipos informáticos (Art. 217); 3.- Vigilancia acústica (Art. 218); 4.- Vigilancia a través de captación de imágenes (Art. 219); y 5.- Vigilancia por seguimiento y localización (Art. 220).

71. Ley Nº 2784, Código Procesal Penal de Neuquén. Este código posee una regulación genérica de la prueba digital, efectuado en tres artículos, ellos son: 1.- comunicaciones (Art 150); 2.- interceptaciones telefónicas (Art 151); 3.- información digital (Art. 153).

72. Ley Nº 5020, Código Procesal Penal de Río Negro. Al igual que el Código de Neuquén posee una regulación genérica de la evidencia electrónica efectuada en a través de tres artículos 1.- Comunicaciones (Art. 145 del CPPRN); 2.- interceptaciones telefónicas (Art. 146 CPPRN), y 3.- incautación de datos (Art. 148 del CPPRN).

73. Ley Nº 8933, Código Procesal Penal de Tucumán, sancionada 20/10/2016. Al igual que el Código de Río Negro reguló la prueba digital en tres artículos. Ellos son: 1.- Comunicaciones (Art. 196 del CPPT); 2.- Apertura y examen de correspondencia. Secuestro. (Art. 197 del CPPT); 3.- Interceptaciones telefónicas (Art. 198 del CPPT); y 4.- Información digital e incautación de datos (Art. 199 del CPPT).

los principios plasmados en El Convenio de Ciberdelincuencia de Budapest (Ley Nº 27411) y con la Ley de Protección de Datos Personales (Ley Nº 25326); sino que resulta indispensable establecer cuáles son los requisitos jurídicos, éticos y técnicos para operar sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal.

Requisitos para la validación legal de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia

Frente al avance exponencial y avasallador de la inteligencia artificial aplicada en el campo del Derecho Penal y del Derecho Procesal Penal, y su implementación en la administración de justicia nacional por medio de sistemas informáticos asistidos por algoritmos controlados por IA, resulta indispensable comenzar a establecer qué requisitos deben cumplir para entender que esos sistemas informáticos asistidos por inteligencia artificial, cumplen con Constitución Nacional, las Convenciones y Tratados de Derechos Humanos, los principios plasmados en El Convenio de Ciberdelincuencia de Budapest (Ley Nº 27411) y con la Ley de Protección de Datos Personales (Ley Nº 25326).

A nivel internacional ya se ha comenzado a trabajar en pautas rectoras para la introducción de inteligencia artificial fiable, siendo Europa un ejemplo de ello.

La Unión Europea, por medio de la Comisión Europea, elaboró "Directrices éticas para una IA fiable", creadas por el "Grupo independiente de expertos de alto nivel sobre inteligencia artificial" en junio de 2018.

De conformidad con las "Directrices éticas para una IA fiable",

> la fiabilidad de la inteligencia artificial (IA) se apoya en tres componentes que deben satisfacer a lo largo de todo el ciclo de vida: a) la IA debe ser lícita, es decir cumplir con todas las leyes y reglamentos aplicables; b) ha de ser ética, de modo que garantice el respeto de los principios y valores éticos; y c) debe ser robusta, tanto desde el punto de vista técnico como social.[74]

74. UE, *Directrices éticas para una IA fiable*, Bruselas, Comisión Europea, 2019, p. 2.

Los tres componentes referidos: 1.- IA legal; 2.- IA ética; y IA robusta, lógicamente comprenden subprincipios rectores.

1. La IA legal implica que los sistemas de inteligencia artificial no operan en un mundo sin leyes. Por el contrario. Esta tecnología debe adaptarse y adecuarse a los derechos humanos receptados por tratados, convenciones y pactos internacionales.[75]
2. La IA ética conlleva que esta tecnología debe respetar los principios de: 1. Respeto a la dignidad humana; 2.- libertad individual; 3.- respeto de la democracia, la justicia y estado de Derecho; 4.- igualdad, no discriminación y solidaridad; 5.-derechos de los ciudadanos.[76]
3. La IA robusta, implica que el sistema debe funcionar de manera segura y fiable.

Habiendo aclarado que toda inteligencia artificial debe presentar una fiabilidad legal, ética y técnica, es menester individualizar qué requisitos específicos debe cumplir un sistema informático asistido por inteligencia artificial (IA) que desee implementarse en la administración de justicia penal, a fin de no resultar contrario a los principios constitucionales o garantías convencionales previstas en los tratados de derechos humanos.

Requisitos para la validación legal de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal

Relevancia de la toma de decisión humana y empleo subsidiario del sistema informático asistido por inteligencia artificial

Los sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal, tanto en el área de "policía predictiva" como de "justicia predictiva", no deben suplir o reemplazar la toma de decisión humana.[77]

75. Ibídem, pp.11-12.

76. Ibídem, pp.11-14.

77. Sprenger, Johanna; Brodowski, Dominik, "'Predictive Policing', 'Predictive Justice', and the use 'Artificial Intelligence' in the Administration of Criminal Justicie in Germany", *Asociación Internacional de Derecho Penal* (*AIDP*), Navarra, Editorial Aranzadi, 2023, p. 33.

Los magistrados, funcionarios, empleados de la administración de justicia, y los miembros de las fuerzas de seguridad, no deben seguir ciegamente y acríticamente los resultados automatizados arrojados por el sistema informático asistido por inteligencia artificial (IA).

La decisión siempre debe depender de un ser humano, sea un juez, fiscal, defensor, oficial de policía, el cual realice una tarea de supervisión de los resultados automatizados arrojados por el sistema informático asistido por inteligencia artificial (IA).

El sistema informático asistido por inteligencia artificial (IA) implementado en el área de "policía predictiva", o bien de "justicia predictiva", no pude constituir el único medio de prueba para la toma de una decisión, y sus resultados deben ser confirmados por otros métodos, lo cual permita mantener la toma de decisión en manos del magistrado o agente de prevención, y no depender en su totalidad de un resultado automatizado brindado por sistema algorítmico asistido por IA.

Ello en consonancia con las "Directrices éticas para una IA fiable", que requieren dentro de una IA robusta, el principio de "acción y supervisión humana", conforme el cual los sistemas de IA deberían respaldar la autonomía y la toma de decisiones de las personas, pero de ninguna manera suplir la toma de decisiones de ellas.

Además, ello se encuentra avalado por el "derecho a no se tomen decisiones basadas únicamente en perfiles realizados a partir de inteligencia artificial".[78] Los procedimientos de selección íntegramente automatizados sin supervisión humana pueden conducir a trato discriminatorio o incluso a falsos positivos del sistema por fallas o sesgos discriminatorios en su programación.

También el principio de relevancia de toma de decisión humana, se encuentra en correspondencia con los preceptos fijados en los precedentes de la jurisprudencia estadounidense "State vs Loomis"[79] y "Loomis vs Wisconsin".[80]

La toma de decisión es un acto sumamente complejo, ya que requiere de sentido común y comprensión del mundo en general, algo

78. Corvalán, Juan, *Perfiles digitales humanos. Proteger datos en la era de la inteligencia artificial. Retos y desafíos del tratamiento automatizado,* Buenos Aires, Editorial Thomson Reuters, pp.137-138.

79. US, "State vs Loomis", Wisconsin, 05/10/2016.

80. US, "Loomis vs Wisconsin", 26/06/2017.

que los sistemas informáticos asistidos por inteligencia artificial, aún no han adquirido.

Téngase en consideración que los seres humanos pueden proporcionar al menos cuatro funciones básicas que la inteligencia artificial aún no posee, ni ha obtenido hasta nuestra actualidad, siendo ellas, capacidad de juicio, empatía con el acusado, creatividad de las soluciones adoptar y versatilidad y adaptabilidad a las circunstancias del caso.[81]

Transparencia algorítmica del sistema informático asistido por inteligencia artificial

Los sistemas informáticos asistidos por inteligencia artificial (IA) implementados en la administración de justicia penal deben contar con lo que se denomina transparencia algorítmica.

La transparencia algorítmica, consiste en que se haya hecho público por parte de la corporación, compañía, empresa privada u organismo público del Estado, que utiliza un sistema informático asistido por inteligencia artificial (IA), la forma en la que funciona el árbol decisional del algoritmo utilizado por el sistema asistido por IA, o bien se haya brindado información clara y precisa de su metodología y funcionamiento.

Está terminantemente vedado utilizar sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal, que operen mediante algoritmos de caja negra (*Black Box*); es decir, aquellos algoritmos en los cuales no está explicitado su metodología o procedimiento, y en los cuales solo se sabe qué datos ingresan y qué resultados surgen.[82]

Resulta indispensable que los sistemas informáticos asistidos por inteligencia artificial (IA) implementados en la administración de justicia penal, posean algoritmos de caja blanca (*White Box*) o transparencia algorítmica.

La demostración de transparencia algorítmica resulta sumamente relevante a la luz del derecho de defensa en juicio, ya que el desconocimiento del diseño del árbol de decisión del *software* de IA, redunda en la

81. Granero Horacio, *GPT-3 y el futuro de la abogacía*, Buenos Aires, Editorial Hammurabi, 2021, p. 207.

82. Castro, Matías, "Juicio a las máquinas", en *Infotechnology*, 2023, pp. 58-60.

imposibilidad del acusado o imputado de defenderse plena y eficientemente mediante el debido control de la prueba de cargo.

Ello en correspondencia con los requisitos de establecidos por la Unión Europea en sus "Directrices éticas para una IA fiable", ya que una IA fiable además de ser legal y ética, requiere ser robusta.

Resulta indispensable que, para ser una IA robusta a nivel técnico, contenga trasparencia algorítmica, lo que involucra: a.- trazabilidad de los datos ingresados; b.- explicabilidad del procedimiento y de los resultados obtenidos; y, c.- comunicación de que dicha decisión fue elaborada por un sistema IA que no involucra una toma de decisión humana.[83]

En igual sentido, este requisito de transparencia algorítmica es demandado a nivel jurisprudencial, por el precedente neerlandés de la Corte de Distrito de la Haya "Rechtbank Den Haag",[84] y también por los requerimientos efectuados por parte de la doctrina.[85]

Por último, y no por ello menos relevante, la trasparencia algorítmica también requiere que los algoritmos implementados en sistemas informáticos asistidos por inteligencia artificial (IA), implementados en la administración de justicia penal, sean sometidos a revisión periódica.

A fin de evitar sesgos de discriminatorios, falsos positivos se recomienda que los algoritmos sean supervisados y sometidos a revisión constante recomendándose a nivel técnico que:

a. Preferir el empleo de *Dataset* dinámicos y no estáticos.
b. Que los sistemas de *Machine Learning* sean entrenados con variedad de data o información y con frecuencia periódica.
c. Testear los sistemas informáticos asistidos con inteligencia artificial con algoritmos de optimización o de fuerza bruta.
d. Los algoritmos deben estar sometidos a revisión en forma constante y permanente, a fin de evitar sesgos discriminatorios o de cualquier tipo a raíz de la desactualización.[86]

83. UE, *Directrices éticas para una IA fiable*, Bruselas, Comisión Europea, 2019, p. 22.

84. CPI, "Rechtbank Den Haag", ECLI:NL: RBDHA, 2020, 865, 05/02/2020.

85. Corvalan, Juan, *Prometea, Inteligencia artificial para transformar organizaciones públicas*, Buenos Aires, Astrea, 2019, p. 60.

86. Castro, Matías, *op. cit.*, pp. 56-61.

Examinación y constatación de la fiabilidad técnica de la metodología empleada árbol decisional del sistema informático asistido por inteligencia artificial

Que la metodología empleada por el árbol de decisión del algoritmo de inteligencia artificial, haya sido previamente revisado por peritos en informática forense por medio de publicaciones en revistas científicas especializadas.[87]

Validación del sistema informático asistido por inteligencia artificial (IA) por parte de la comunidad científica especializada en implementación técnica de IA

La validación del sistema informático asistido por inteligencia artificial (IA) en la administración de justicia penal debería ser otorgada por parte de la comunidad científica especializada en informática forense, ya sea a través de proyectos de investigación presentados en ámbitos académicos y/o centros internacionales investigación destinados al estudio específico de esta área.

Ello resulta indispensable dado que todo programa forense asistido por inteligencia artificial para ser fiable debe poseer un mecanismo de solidez técnica[88] que involucra su sometimiento a pruebas y testeos por parte de la comunidad científica.

Creación dentro del Estado Nacional de una Comisión Científica Evaluadora de Algoritmos de Inteligencia Artificial (IA)

Como propuesta de *lege ferenda* se recomienda que el Estado Nacional cree organismos públicos destinados a auditar los códigos fuentes, árboles decisionales o algoritmos empleados por los sistemas informáticos asistidos por inteligencia artificial (IA) destinados a ser empleados en la administración de justicia penal;[89] o bien, que el Estado Nacional cree una única Comisión Científica Evaluadora de Algoritmos de Inteli-

87. Polansky, Jonathan, *Garantías constitucionales del procedimiento penal en el entorno digital.* Buenos Aires, Editorial Hammurabi, 2020, p. 148.

88. UE, Directrices éticas para una IA fiable, Bruselas, Comisión Europea, 2019, pp. 20-21.

89. Ibídem, p. 24.

gencia Artificial (IA)[90] destinada a evaluar y auditar los algoritmos de los sistemas informáticos asistidos por inteligencia artificial (IA) destinados a ser utilizados en la administración de justicia criminal.

Conclusión

Los sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal de la República Argentina ya se encuentran presentes desde fines de la década pasada, tanto en el área de la "policía predictiva" como la "justicia predictiva".

Sistemas como Sherlock Legal y *Prometea* en el área de la "justicia predictiva", y el Sistema de Reconocimiento Facial de Prófugos (SRFP) de la Ciudad Autónoma de Buenos Aires (CABA), y otros sistemas de reconocimiento facial en las provincias de Buenos Aires, Córdoba, Mendoza, Salta y Santa Fe, en el área de "policía predictiva", son un claro testimonio de la clara irrupción de la inteligencia artificial en la justicia penal.

Los acelerados avances en inteligencia artificial solo harán profundizar este escenario, incrementando en esta década (2020-2030) los sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal, probablemente a través de asistentes jurídicos inteligentes, sistemas informáticos asistidos por IA destinados al análisis computacional del derecho; la elaboración automatizada de proyectos de sentencias se tornará cada vez más cotidiana.

Desgraciadamente, hasta la fecha el ordenamiento jurídico argentino no cuenta con ninguna ley o norma dentro de los códigos procesales penales nacionales, federales o provinciales que defina la inteligencia artificial, los sistemas informáticos asistidos por inteligencia artificial, policía predictiva, o justicia predictiva.

Debido a que en la era de la cuarta revolución industrial (4.0) la inteligencia artificial pareciera estar llamada a procesar volúmenes astronómicos de información, los cuales resultan imposibles de analizar por la mente humana, dando lugar a un nuevo régimen en torno a la verdad, pareciendo destronar a la especie humana como productora del saber y como sujeto cognoscente; resulta indispensable establecer

90. Polansky Jonathan, *op. cit.*, p. 103.

requisitos y parámetros para la implementación legal de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal.

Resulta claro que la incorporación de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal, debe cumplimentar con requisitos tales como:

1. Relevancia de la toma de decisión humana y empleo subsidiario del sistema informático asistido por inteligencia artificial solo como una herramienta político-criminal.
2. Transparencia algorítmica del sistema informático asistido por inteligencia artificial.
3. Examinación y constatación de la fiabilidad técnica de la metodología empleada árbol decisional del sistema informático asistido por inteligencia artificial.
4. Validación del sistema informático asistido por inteligencia artificial (IA) por parte de la comunidad científica especializada en implementación técnica de IA.
5. Creación dentro del Estado Nacional de una Comisión Científica Evaluadora de Algoritmos de Inteligencia Artificial (IA).

Estos principios rectores, y otros que probablemente estén próximos a surgir, resultaran indispensables para poder afirmar que el empleo de sistemas informáticos asistidos por inteligencia artificial (IA) en la administración de justicia penal en la República Argentina resultan respetuosos de la Constitución Nacional, las Convenciones y Tratados de Derechos Humanos, los principios plasmados en El Convenio de Ciberdelincuencia de Budapest (Ley Nº 27411) y con la Ley de Protección de Datos Personales (Ley Nº 25326).

Inteligencia Artificial y derecho penal: realidades y proyecciones*

Marcelo A. Riquert**

Notas introductorias

Asistimos en la actualidad a lo que muchos definen como el comienzo de la cuarta revolución industrial, signada por el surgimiento o avances significativos en áreas tales como la nanotecnología, la biotecnología, el *cloud computing*, la impresión en 3D, la *Big Data* y, por supuesto, la inteligencia artificial (en adelante, IA). La aparición de la máquina a vapor había signado a la primera, en el siglo XVIII; la electricidad y el fordismo moldearon la segunda en el siglo XIX, mientras que el siglo pasado fue atravesado por dos guerras mundiales entre las que hubo la gran depresión y, en lo tecnológico, ofreció, entre otras cosas, la energía nuclear (que dio por terminada la segunda gran guerra y fue factor decisivo para la instalación de llamada "guerra fría") y el geométrico crecimiento de las llamadas tecnologías de la información y comunicación (TIC). Así la tercera revolución industrial marcó sus últimas décadas signadas por la "globalización". Para el derecho penal esta fenomenología llevó a una larga discusión acerca de la posibilidad o no de

* El texto corresponde a la ponencia presentada en el "Coloquio Internacional: Inteligencia artificial y administración de justicia, la policía y la justicia predictiva", organizado por el Grupo Argentino de la AIDP y celebrado en el Salón Azul de la Facultad de Derecho de la Universidad de Buenos Aires, del 28 al 31 de marzo de 2023. La exposición integró la Mesa Redonda "IA en la actualidad Argentina", realizada el 29 de marzo e integrada además por los profesores Nora Cherñavsky y Carlos Christian Sueiro. Se ha mantenido en lo posible la exposición original, agregando al final un listado de la principal bibliografía consultada para su elaboración. El desarrollo en extenso se ha concretado en la obra *Inteligencia artificial y derecho penal*, prologada por E. Raúl Zaffaroni, CABA, Ediar, 2022.

** Director del Área Departamental Penal de la Facultad de Derecho de la Universidad Nacional de Mar del Plata (Argentina). Ex Presidente de la Asociación Argentina de Profesores de Derecho Penal. Miembro del Grupo Argentino de la AIDP.

procesar la conflictividad inherente con las herramientas del derecho penal clásico o "nuclear" (así, la escuela de Frankfurt) o si resultaba necesario un nuevo derecho penal cuando ya no se trata simplemente de bienes jurídicos individuales sino macrosociales, autores que no sólo son personas físicas sino también jurídicas y técnicas de tipificación en que el resultado no es un daño (debemos prevenirlo) sino un peligro.

Mientras en la tercera revolución industrial, frente a las conductas disvaliosas realizadas por medio de las TIC nos preguntábamos si había "delitos informáticos" o, por su influencia como factor criminógeno, "delincuencia informática"; la cuarta, como no puede ser de otro modo, nuevamente detrás del cambio tecnológico, nos encuentra planteándonos qué hacer con lo que sucede en el "metaverso" (nuevo espacio digital inmersivo e interactivo para la conflictividad humana) y si regulamos y cómo lo hacemos a la IA, justamente objeto de esta reunión. Cuando se sostiene el pasaje del "antropoceno" al "tecnoceno" (tal como refiere Flavia Costa), estamos viviendo en la que se define como la "sociedad de la información" y si nos preguntamos por cuánta información hay acumulada el aserto se revela acertado cuando el volumen se tiene que medir en zetabytes (un uno seguido de veintiún ceros). Dos datos que por "viejos" no pierden elocuencia: 1) entre 2014 y 2017 lo digitalizado equivale a todo lo acumulado desde la prehistoria hasta 2014; 2) hasta 2015 se habían reunido 5 zetabytes que, visibilizados en papel, equivalen a 4500 pilas de libros hasta el sol. La tecnología sigue esparciéndose por el mundo a ritmo acelerado. Desde 2020 la población del planeta creció un 4% (pasamos de 7700 millones a 8000 millones de personas), lapso en que la conectividad con Internet lo hizo el 8% llegando a un total de unos 5100 millones de individuos (68% del total mundial) y el uso de redes sociales subió un 9,4%, interactuando en ellas unos 4760 millones de usuarios (59,4% del mismo total). Por supuesto, esto se da con marcadas asimetrías cuyos extremos eran en 2020 Emiratos Árabes Unidos, con el 99% de su población conectada, y Burundi, con solo el 7%. Siendo las cifras de crecimiento impactantes y, aunque en ámbitos como en el que ahora mismo estamos reunidos, podemos sentir que todos participamos del mismo fenómeno, lo cierto es que todavía el 40% de los terrestres no viven su vida en contacto con las redes sociales, fuera de sus beneficios y también de su toxicidad; y el 30% no tiene acceso permanente, estable, a Internet.

De revés, el 70%, en forma más o menos directa, estamos viviendo lo que Zuboff denomina como la era del capitalismo de la vigilancia, la Internet de las cosas y el mercado de los futuros conductuales, y que Rouvroy y Berns llaman la "gubernamentalidad algorítmica". Pasamos de la inteligencia de las máquinas a casas inteligentes donde todo lo que hacemos se registra por nuestros proveedores, lo que permite no sólo predecir sino también orientar nuestra conducta futura pero no en nuestro propio beneficio sino el de otros, profundizando el instrumentarismo. Y los proveedores, tanto en occidente como en oriente, son un acotadísimo número de jugadores que Amy Webb definió como los "Big 9": de un lado Google, Microsoft, Apple, Facebook, IBM y Amazon (GMAFIA); del otro, Baidú, Alibaba y Tencent (BAT). Otros ofrecen mínimos cambios y sintetizan en GAM(F)A (Google, Apple, Meta –Facebook– y Amazon) y BATX (Baidú, Alibaba, Tencent y Xiaomi). Si bien una mirada sesgada y superficial parece inquietar a occidente ante el "programa del buen ciudadano" de China, simplificando, la realidad es que la operatividad de los grupos no ofrece diferencias significativas porque de este lado del mundo también la información algorítmicamente procesada determina si nos dan o no un crédito, si accedemos o no a un trabajo, si somos objeto de cancelación social o no. En fin, lo que llevó a que Cathy O'Neil se refiriera a los algoritmos como "máquinas de destrucción matemática".

Podríamos advertir que una diferencia palpable comienza a mostrar Europa, primero con su "Carta Ética Europea sobre el uso de la IA en los sistemas judiciales y su entorno" (2018), más su Res. de 6/10/21,[1] y luego, de mayor amplitud, cuando el año pasado promovió leyes para la transparencia digital, como la Ley de Servicios Digitales y la Ley de Mercados Digitales. Se procura con ellas que se alcancen estándares mínimos de transparencia digital por las tecnológicas con más de 45.000.000 de usuarios, es decir, las que importan: Google, Meta, Instagram, Twitter, Amazon.

Los ostensibles progresos en el desarrollo de dispositivos de interfaz entre cerebro y computadora, tanto externos como invasivos, junto a los descubrimientos en materia de neurociencias, detrás de lo que

1. Res. del Parlamento Europeo de 6 de octubre de 2021 sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y judiciales en asuntos penales.

está tanto la empresa privada (por ejemplo, Neurolink, de Elon Musk, o Facebook) o los estados (China), han provocado que comencemos a hablar de "neuroderechos". En el proyecto "Brain" de Rafael Yuste (Universidad de Columbia) se identifican cinco: privacidad mental, identidad personal, libre albedrío, aumento de la neurocognición y protección respecto de los sesgos. Otro integrante de este mismo panel, el Dr. Sueiro, ha mencionado la discusión en torno a la necesidad de futuras tipificaciones de conductas como el acceso ilegítimo a la mente de una persona por medio de un dispositivo de interface cerebro-computadora (ICC) o el apoderamiento de información mediante un dispositivo de ICC. Es importante no perder de vista todo este contexto y, retomando la ineludible nota de una regulación jurídica que va por detrás de los acontecimientos, parece oportuno señalar la coincidencia con la propuesta de Haissiner y Pastor: "frente la tecnología, ni fobia ni euforia".

Inteligencia artificial y su consideración jurídica

En la pasada década de los noventa Ulrich Sieber nos decía que los avances en las tecnologías de la información y comunicación habían ofrecido como consecuencia cuatro oleadas de reforma legal:

1. la protección de la privacidad en los setenta;
2. la represión de los delitos económicos por ordenadores en los ochenta;
3. contemporánea, la protección de la propiedad intelectual en la informática; y,
4. la reforma procesal para el tratamiento de la evidencia digital en los noventa.

Hemos agregado una quinta oleada a comienzos del actual milenio de la mano de la instalada lucha contra el enemigo terrorista derivada de los ataques del 9/11, popularizándose la discusión en torno al hacktivismo y el ciberterrorismo. Poco después, los avances en materia de IA y el avance de la Internet de las cosas nos enfrentan a nuevos desafíos jurídicos y crece la consideración de una suerte de sexta ola que comienza a tener reflejo en los estudios de grado, observándose en los más recientes programas de estudios de la carrera de abogacía

que se incorporan materias como "IA y derecho" o "Derecho Artificial" (denominación la última que no me parece muy feliz).

Ha pasado mucho desde aquella inicial reunión en Dartmouth (1956) para responder al interrogante: ¿es posible que una máquina simule la inteligencia humana?[2]. La imagen de los científicos de distintas áreas del conocimiento reunidos bajo aquella premisa basta para contestar otra pregunta de suma pertinencia como es si los algoritmos son asépticos o ideologizados, alternativa que decanta con claridad hacia lo último. La fotografía muestra palmarias subrepresentaciones en lo racial, etario y género. Podría contra-argumentarse que hoy no sería exactamente igual y es cierto, no lo sería porque habría algunos cambios en los porcentajes de representación pero no serían realmente significativos en términos de pertenencia cultural.

Si tomamos como válida la definición que se proporcionara en los cuestionarios recibidos de AIDP para producir los informes nacionales, se entiende por IA un sistema que muestre un comportamiento inteligente al analizar su entorno y adopte cursos de acción, con cierto grado de autonomía, para lograr objetivos específicos (Picotti). Para entender dónde estamos parados, respecto de qué discutimos cuál debe ser la consideración jurídica, es conveniente recordar que se suele distinguir entre la IA débil o estrecha, la promedio o general y la fuerte o súper IA. La débil refiere a los sistemas informáticos que permiten que el aprendizaje automático realice una tarea específica. La promedio a los que tienen capacidad de comprensión para cualquier tarea. La fuerte a los que exceden la capacidad de los seres humanos. ¿Dónde estamos parados hoy? Según se puede leer con consenso, todavía en la débil o estrecha. Tal limitación, valga la insistencia, no elude el problema general de ser útil para amplificar, potenciar y replicar rasgos negativos. Valga como ejemplo el uso de algoritmos para difundir y multiplicar los discursos de odio, de inquietante actualidad. ¿Y qué se está haciendo? Desde su sede de París, la Organización para la Cooperación y el Desarrollo Económico (OCDE), que reúne en la actualidad a 42 naciones, entre

2. Me refiero a la conocida como "Conferencia de Dartmouth" por haberse celebrado en el Dartmouth College, en Hanover, New Hampshire, Estados Unidos, durante el verano de 1956, tras una propuesta en setiembre de 1955 elaborada por John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon. Duró 6 semanas y tuvo un total de 11 participantes.

ellas, Argentina, ha creado un Consejo sobre IA. Se discuten dos modelos para regular la IA: 1) normas verticales, sectoriales y específicas; 2) regulación transversal por tipo de tecnología. También dos modelos de control: a) un órgano guardián de IA (por ejemplo, en Argentina hay un proyecto de ley en trámite para un Consejo Federal de IA); b) consejos de seguridad sectoriales. Los problemas que deben ser tenidos en cuenta: a) la caja negra (transparencia de los sistemas); b) los sesgos en los algoritmos; c) la ética de la selección (¿es lo mismo perseguir el terrorismo que enfrentar una crisis sanitaria como la pandemia por el COVID-19?); d) el manejo de la información.

En lo que hace específicamente a lo penal, es de interés mencionar otra vez a la "Res. del Parlamento Europeo de 6 de octubre de 2021 sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y judiciales en asuntos penales". Allí se resalta que la IA ofrece tanto posibilidades como riesgos extraordinarios, que no es un fin en sí misma sino un instrumento al servicio de las personas. Se enfatiza que el objetivo debe ser aumentar el bienestar, capacidad y seguridad humanos. Y advierte que su uso en el ámbito del derecho penal se basa en la promesa no siempre cumplida de reducir la criminalidad. No hay naturalmente espacio en esta ponencia de explorar la totalidad del universo de alternativas que ofrece la IA al presente y, con la velocidad que se suceden las novedades, ciertamente es casi imposible estar actualizado. Entre las principales ramas de la IA suele mencionarse al *machine learning*, el diseño de sistemas expertos, la visión artificial, el reconocimiento del habla, la planificación automática, la robótica y el procesamiento de lenguaje natural. En esta última proliferan las noticias a partir del impacto que Open IA ha logrado a fines del año pasado y comienzos del presente con su "ChatGPT" (Chat Generativo Preentrenado Transformador). Hace unos días los diarios publicaban que un juez colombiano había dictado una sentencia asistido por el ChatGPT. Con lógica curiosidad, comencé a experimentar con esta herramienta que el Dr. Ingeniero Informático José Ángel Olivas Varela (UCLM-Ciudad Real) caracterizó recientemente como un "Google anabolizado", lo hice planteándole la necesidad de hacer un escrito de eximición de prisión. Brindé los datos básicos y en segundos tuve la propuesta de escrito en el que se habían incorporado algunos argumentos pertinentes que yo no había proporcionado. El ChatGPT había consignado que el

imputado no tenía antecedentes penales y que tenía arraigo domiciliario y familiar que le servirían de contención. Así, por un lado, reveló que tenía en cuenta factores de interés para la decisión que se solicitaba; por el otro, "inventó" datos de los que, en realidad, no disponía. Déficit manifiesto fue la falta de indicación normativa para fundar el pedido. A través de varias correcciones, finalmente obtuve un escrito en condiciones de ser presentado. Mi conclusión personal sintética: mejor que mi expectativa inicial, pero lejos todavía de ofrecer un escrito listo para usarse (¿por cuánto tiempo?, sería la pregunta subsiguiente).

Además, le pedí que me ayudara a preparar esta conferencia interrogándole sobre la utilidad de la IA para cometer delitos. La reconoció y mencionó como ejemplo de una aplicación maliciosa de la IA la creación de programas de phishing o malware. Enumero varios posibles delitos en que sería de ayuda: fraude financiero, ataques informáticos, espionaje y vigilancia ilegal, abusos online y discursos de odio (sobre los últimos volveré después).

También aclaró que la IA no es ni buena ni mala sino una herramienta neutral que puede ser usada para fines positivos o negativos dependiendo la voluntad del usuario. En el primer sentido ejemplificó que la IA puede ser utilizada para detectar delitos, para mejorar la seguridad de los sistemas y la privacidad online.

En términos del posible uso para detectar delitos no puedo dejar de señalar que, en ocasiones, puede ofrecer una nueva gama de problemas desde el punto de vista penal. Así, la Universidad de Deusto ha desarrollado un negobot (Lolita), que opera cual suerte de agente provocador online para detectar pedófilos y la distribución de material de abuso sexual infantil. "Lolita" interactúa simulando ser una niña con los potenciales ofensores sexuales, que creen estar ante una joven real a la que procuran engañar para perpetrar su criminal designio. La utilidad, indudable. Que acarrea todos los problemas y discusiones inherentes al uso y límites de la provocación, también indudable.

Si retomamos el tema de los discursos de odio, la IA se suele asociar tanto como un factor potenciador negativo, facilitando su reproducción y multiplicación, como una herramienta fundamental para su detección y prevención. En marzo de 2023, en el seminario internacional "Inteligencia Artificial y Sistema Penal", celebrado en la Facultad de Derecho y Ciencias Sociales de Ciudad Real, Universidad de Castilla-La

Mancha, coorganizado por su Escuela de Ingeniería Informática y el Instituto de Derecho Penal Europeo e Internacional, se mostraron los avances un proyecto conjunto entre ambas áreas para la detección y evaluación (valoración jurídica) de estos discursos online[3], bajo apropiados estándares de transparencia[4]. Retomando la consulta con el ChatGPT, justamente indicó como opciones preventivas que una IA puede desarrollar en la prevención del discurso de odio a la detección automática de contenidos ofensivos, su monitorización en tiempo real, el análisis de tendencias y patrones en la publicación y, finalmente, la generación de contenidos alternativos que promuevan la tolerancia y la inclusión. No es esta la ocasión para profundizar sobre esta temática pero sí luce oportuna su mención cuando se habla sobre si regular y cómo hacerlo lo vinculado con la IA, en particular porque se trata de una temática en que la tensión entre la libertad de expresión y los límites generales de prohibición han mostrado cómo la adopción de modelos contrapuestos ha entorpecido el avance en la materia. En efecto, pueden observarse en nuestra esfera de referencia cultural, es decir, occidente, que cohabitan dos modelos para regular los discursos de odio y las *fake news*: de un lado el estadounidense, que propicia al máximo la libertad de expresión y trabaja bajo premisa de inexistencia de censura previa y responsabilidad posterior por daño (sería el que sigue Argentina conforme dejan en claro diversos precedentes de la CSJN como "Rodríguez", "Gimbutas", "Paquez" y, el último, "Denegri" –cuando se trató la cuestión del derecho al olvido–); del otro, el que procura evitar la producción del daño y, por lo tanto, implementa la censura previa, lo que se ha visto en forma restringida a cuestiones electorales en reciente normativa francesa y alemana. En síntesis, que el ejemplo es demostrativo que no basta con acordar sobre la necesidad de regular, sino que lo más difícil es acercar posiciones acerca de cómo hacerlo.

3. Valga recordar sobre el particular la importancia del "Plan de Acción de Rabat" (ACNUDH), que propende se haga la valoración siguiendo parámetros vinculados al contexto, orador, extensión, intención, probabilidad/inminencia y contenido/formas.

4. Destaco en este sentido la presentación del doctorando Andrés Montoro Montarroso, ya que el desarrollo es parte de su tesis, que avanza dirigida por los Catedráticos Dres. José Angel Olivas Varela (Informática) y Adán Nieto Martín (Derecho).

IA y derecho penal: problemas fondales y procesales

Yendo a la específica vinculación entre la IA y el derecho penal, puede decirse que media cierto grado de consenso en que se han detectado problemas tanto en lo atinente a la rama sustantiva como la adjetiva del último. Las áreas en que se focalizan puede sistematizarse del siguiente modo: a) tecnología de predicción criminal (aquí, por ejemplo, lo relativo a la vigilancia predictiva y los sistemas de reconocimiento facial, por voz, etc.); b) agentes autónomos o artificiales (hay numerosas cuestiones relativas a la autoría y participación criminal por los resultados disvaliosos cometidos por autos de conducción autónoma y drones tanto de uso civil como militar); c) pandemia y control sanitario (el COVID-19 dio lugar a un uso intensivo de la IA tanto para la prevención y tratamiento, como para la instalación de la cibervigilancia); d) gestión judicial (en particular, el apoyo de decisiones sobre la base de la valoración algorítmica de riesgos). Por tomar solo un ejemplo, baste tener presente la extensión que ha alcanzado la vigilancia tecnológica del espacio público (y sus posibilidades de avance en el privado), siendo probablemente extremos la desarrollada en las grandes ciudades chinas por intermedio de la compañía SenseTime o la proliferación de cámaras londinense. Harari ha señalado[5] que la cibervigilancia se presentó para la pandemia del coronavirus como una suerte de "solucionismo tecnológico" que importó el pasaje de la que calificó como "epidérmica" (*Big Data* sobre datos sensibles y geolocalización) hacia la que llama "hipodérmica" (por su profundización a partir del uso de los datos biométricos). Resaltó que en aquel contexto la IA sirvió tanto para combatir la infodemia y prevenir y contener contagios, como para instalar una suerte de lo que podría llamarse "coronoptikón" (así, Ramonet), un panóptico a propósito del coronavirus. No puede soslayarse que, en la actualidad, en los países más desarrollados pero también en los que no lo son –aunque probablemente con menos intensidad y calidad de medios– ya se usan

5. Harari, Yuval, "El mundo después del coronavirus", trad. por Juan Gabriel López Guix, en *La Vanguardia*. Disponible en: https://www.lavanguardia.com/internacional/20200405/48285133216/yuval-harari-mundo-despues-coronavirus.html [fecha de consulta: 18/09/2024].

o están disponibles numerosas "apps" de IA para el uso policial. Sin pretensión de exhaustividad pueden mencionarse:

a. Reconocimiento facial (en Argentina, implementada por ejemplo en CABA y Mendoza);
b. Reconocimiento de matrículas (ídem, en CABA)
c. Identificación por voz y reconocimiento del habla (en uso por el MPF de la CABA, con el sistema "Prometea");
d. Reconocimiento de temperatura (por el Covid, múltiples puntos de uso en estaciones de transporte público, aeropuertos, grandes centros comerciales, etc.);
e. Tecnología de lectura de labios y vigilancia auditiva (algunos sistemas de cámaras de vigilancia del espacio público en CABA tienen posibilidad de hacer tomas de sonidos por un tiempo limitado);
f. Análisis autónomo de bases de datos no identificadas;
g. Predicción de puntos críticos delictivos (*hot spots*)
h. Detección de comportamientos (en la localidad bonaerense de Tigre, hay un sistema que detecta patrones de recorrido);
i. Herramientas de autopsia virtual;
j. Herramientas para detección de fraudes financieros y financiación del terrorismo;
k. Forensia sobre dispositivos electrónicos: UFED (*Universal Forensic Extraction Device*).

¿Cuáles son los riesgos de estas herramientas? En general, se suele mencionar que, al menos todavía, exhiben variables grados de fiabilidad y precisión lo que traduce en resultados con distintos porcentajes de errores o falsos positivos. Pero no es todo y, quizá, tampoco lo más importante. En efecto, no hay dudas que en poco tiempo mejorarán y los yerros serán mínimos. El problema más significativo es y seguirá siendo la repercusión sobre los derechos fundamentales, lo que abarca un amplio prisma en el que se incluyen desde los sesgos y la discriminación, hasta la tutela judicial efectiva, el juez imparcial, la libertad de expresión, el principio de inocencia, los bucles de criminalización o la opacidad de decisiones, entre otras cuestiones.

Las voces de alerta ya han tenido variada recepción. Por caso, la referida "Carta Ética europea…" de 2018, indica la necesidad de respetar los principios de respeto de los derechos humanos; calidad y segu-

ridad; control del usuario; transparencia, imparcialidad y equidad; y no discriminación. En nuestro país, en setiembre de 2022, en el marco de un Congreso en el que se dedicó una Comisión al tema del impacto de la IA en el derecho civil,[6] se aconsejó en las conclusiones que la regulación nacional debía hacerse bajo criterios bastante similares: abordaje multi y transdisciplinar; contextualizar el fenómeno en el respeto y defensa de los derechos humanos; atender al pluralismo; respetar la dignidad de la persona; respetar los estándares de recolección, transferencia y tratamiento de datos de perfilamiento; y protección del perfil digital humano.

La mencionada "Resolución del Parlamento Europeo de 6 de octubre de 2021 sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y judiciales en asuntos penales" ha señalado que los sistemas de IA que se usen deben ser fiables y que tal estándar debe verificarse aunque la información provenga de apps de IA aprobadas en terceros países; que las apps deben ser plurales desde el diseño (esto quiere decir, no discriminatorias, seguras, transparentes, explicables y respetar la autonomía humana y los derechos fundamentales); y, finalmente, que tanto la policía como la justicia deben usar apps con respeto del mismo estándar, de lo que resulta garante el Estado.

Provisorias conclusiones

Para cerrar, me parece oportuno insistir en aquello de "ni fobia, ni euforia" y señalar que es central no perder de vista que el derecho penal interviene, con límites, después del delito. En este sentido, la mencionada Resolución Europea llama la atención sobre la necesidad que el uso de la IA en materia penal no se aparte de principios básicos que ha costado enorme esfuerzo sentar como básicos en los modernos estados de derecho. Podría pensarse que tal vez no fuera preciso al momento actual de nuestro desarrollo como sociedades. Sin embargo, participo de la idea de que nunca está de más recordarlos, evitar que

6. Me refiero a las XXVIII Jornadas Nacionales de Derecho Civil, celebradas en Mendoza del 22 al 24 de setiembre de 2022, durante las que la Comisión Nº 10 se dedicó a "Transdisciplina, Inteligencia Artificial, mercado y ética", bajo presidencia de los Dres. Andía y Molina Quiroga y vicepresidencia de los Dres. Adaro, Corvalán y Santarelli.

se pierdan de vista. No puede negarse que es probable que algunos de ellos sufran modificaciones (que pretendemos sean las mínimas posibles), que no permanezcan inalterados, alguna mella, algún esmerilamiento habrá. Y ello sería tolerable en la medida que no afecte su esencia, que no pierdan centralidad.

En fin, que la regulación que logremos debe observar respeto por las ideas de que no puede consagrarse un derecho penal de autor, que negamos el avance del estado de policía, que decimos no al peligrosismo y que aún cuando fuera técnicamente viable no puede habilitarse una vigilancia masiva ilimitada. En síntesis, que la razón de eficacia no legitima el olvido del sistema de garantías penales construido sobre siglos de sangre. Siendo esto así, no albergo dudas en torno a que las apps de IA que colisionen con los derechos fundamentales deben directamente ser prohibidas.

Sistemas automatizados. Sistemas predictivos. Inteligencia Artificial y *Big Data*. Decisiones basadas en datos. Modelos de aplicación en el ámbito jurídico

Nora A. Cherñavsky*

Introducción

Desde el año 2006 quienes venimos investigando el cibercrimen y asesorando en el desarrollo de legislación, reglamentos, protocolos y tratados multilaterales –como el Convenio de Budapest–, y que hemos participado del proceso de discusión de la Ley Nº 26388 –que finalmente adapta los tipos penales tradicionales de acceso, daño y fraude entre otros, al medio digital–, venimos participando también en una antigua discusión que se remonta a la edad media respecto a la responsabilidad de las corporaciones, remozada en el siglo pasado en los años 60 y luego desarrollada en los 80, en torno a la "Responsabilidad Penal de las Personas Jurídicas".

Esta última discusión, más anclada en los actuales desarrollos tecnológicos, gira en torno de la responsabilidad y el título de la imputación a los ISP (Proveedores de Servicio de Internet).

Lo más importante en este debate sobre regulación y autorregulación tiene que ver fundamentalmente con el uso de nuestros datos: los ISP almacenan pentabytes de datos de todos nosotros, los que son objeto de tratamiento, entrecruzamiento, segmentación y perfilamiento; por lo tanto, el punto de conexión con su responsabilidad civil, administrativa y penal tiene que ver con que las ganancias se las quedan justamente estos intermediarios de servicios de red.

* Profesora Regular Adjunta de Derecho Penal y Procesal Penal de la Facultad de Derecho de la UBA.

La cuestión de la circulación de contenidos vinculados a la intimidad, específicamente a la imagen de las personas, fue discutida por la Corte Suprema de Justicia de la Nación en "Ponzetti de Balbin c/Editorial Atlántida",[1] antes de la era de Internet y más recientemente en 2014 en el precedente conocido como "Rodriguez María Belen c/Google",[2] en el que la Corte deja sentado el principio de libertad de expresión respecto de los contenidos que circulan por Internet y, por otra parte, asienta la responsabilidad subjetiva del Proveedor de Servicios por contenidos de terceros, esto es, que una vez anoticiado del contenido infractor (*take notice*), el intermediario debe proceder a darlo de baja en forma diligente (*take down*).

Dicho contenido debe ser individualizado mediante el URL[3] por la persona damnificada y/o por el Juez de la causa, en su caso, para que nazca la obligación del proveedor de retirada del contenido infractor.

En este sentido, el debate se centra en la necesidad de pasar de un sistema de autorregulación y aceptación de los términos generales del servicio, y de su contratación por parte de los usuarios-consumidores de Internet, a un sistema de regulación de las plataformas digitales,[4] en tanto que ellas estructuran la comunicación y son los actores más importantes del ecosistema de Internet, ya que nos proveen de acceso y conectividad, alojamiento y contenidos, aplicaciones en la nube y otros servicios esenciales de la sociedad de la información.

1. CSJN, "Indalia Ponzetti de Balbin c/ Editorial Atlántida S.A", 11/12/1984.

2. CSJN, "Rodriguez María Belén c/Google Inc y otro s/daños y perjuicios", 28/10/2014.

3. URL, del inglés Uniform Resource Locator, es un identificador de recursos uniforme. Están formados por una secuencia de caracteres, de acuerdo a un formato modélico y estándar, que designa recursos en una red, como Internet (localizador universal de recursos). El URL es una cadena de caracteres con la cual se asigna una dirección única a cada uno de los recursos de información disponibles en Internet. Existe un URL único para cada página de cada uno de los documentos de la WWW, para todos los elementos de Gopher y todos los grupos de debate Usenet, y así sucesivamente. El URL de un recurso de información es su dirección en Internet, la cual permite que el navegador web la encuentre y la muestre de forma adecuada. Por ello, el URL combina el nombre de la computadora que proporciona la información, el directorio donde se encuentra, el nombre del archivo, y el protocolo a usar para recuperar los datos para que no se pierda alguna información sobre dicho factor.

4. Disponible en: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act/europe-fit-digital-age-new-online-rules-platforms_es

La regulación europea precisamente pretende introducir entre otras medidas de protección al usuario las de transparencia, trazabilidad de vendedores en las plataformas de intercambio de servicios y mayores garantías para los usuarios de poder impugnar decisiones de moderadores de las plataformas y medidas de protección contra los contenidos delictivos en línea.[5]

La Inteligencia Artificial y el *Big Data*

Transcurridas ya dos décadas del siglo XXI tomamos conciencia que tecnologías disruptivas como la inteligencia artificial y la biotecnología están ofreciendo a la humanidad el poder de remodelar y rediseñar la vida en forma radical.

A su vez, somos conscientes que, más allá de las discusiones filosóficas,[6] es a los juristas a quienes nos toca analizar los cambios y si –y en qué medida– las personas inmateriales (una IA) pueden ser sujeto del derecho penal y en dicho supuesto determinar el título de su imputación; si dicha responsabilidad, sea civil, administrativa o penal es trasladable a sus desarrolladores y/o a quienes las pusieron en el mercado negligentemente, es decir, sin la debida supervisión y control de riesgos.

Este análisis resulta pertinente con el desarrollo de la robótica y de los sistemas asistidos por Inteligencia Artificial que ya envuelven muchos de los ámbitos de nuestra actuación, incluyendo el ámbito de la justicia.

Voy a detenerme en estas tecnologías basadas en datos, fenómeno conocido como de *Big Data*,[7] las que ya han comenzado a generar enormes

5. Disponible en: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-platforms_es [fecha de consulta: 06/03/2024]

6. Disponible en: https://archive.is/xCyKU [fecha de consulta: 06/03/2024]

7. *Big data* se refiere a una gran cantidad de información que sólo se puede procesar mediante el uso de herramientas digitales y que sirve para responder preguntas a través del análisis de enormes volúmenes de datos. Se trata de un paradigma se caracteriza por lo que se conoce como las cinco V del *Big Data*: volumen, velocidad, variedad, veracidad y valor. Volumen: Gran cantidad de datos procesados con herramientas digitales, Velocidad: Transmisión de enormes cantidades de datos de forma continua. Variedad:

cambios[8] y paralelamente nuevos riesgos, en forma similar a lo que en el siglo XVIII sucedió con los desarrollos de la revolución industrial.

Hoy asistimos a una transformación aún más radical, en el que la IA promete cambiar de plano no sólo las formas de adquirir conocimiento, sino la economía mundial.

Los sistemas basados en macro datos prometen no sólo suplementar y acrecentar las capacidades humanas, sino también sustituir sus decisiones, presentándose a su vez como sistemas más objetivos y justos para la toma de decisiones.

El desarrollo de los Chats GPT[9] de Open AI de Microsoft fue posible gracias al análisis y a la explotación de la cantidad exponencial de datos brindados por todos nosotros, sin que todavía se haya resuelto la cuestión de la propiedad de los datos, en tanto producto mercantilizable.[10]

Este ChatBot, "va adquiriendo el dominio del lenguaje"[11] y reelaborando la realidad, seleccionando imágenes y contenidos, y escribiendo poesías, además promete el cambio en la economía mundial, y predice grandes beneficios y progresos para la actividad humana, que a la vez supone innumerables desafíos y costos que deberán recaer en "la sociedad en su conjunto". Es fácil advertir que esos costos serán sin duda, mayores para algunos que para otros, como así también los riesgos, si es que no se regula su distribución.

Esta primera cuestión, acerca de quién solventará los nuevos costos sociales y laborales por la incorporación de Copilot en Office 365, el cual introduce la inteligencia artificial en la vida cotidiana de millones de trabajadores, mejorando la productividad y ahorrando tiempo en tareas rutinarias, resulta paralela a la que nos formulamos en el ámbito jurídico penal acerca de quién será responsable por los errores o

Datos recolectadas de diversas fuentes de información. Veracidad: Fuentes de información confiables para su análisis. Valor: Creación de nuevas oportunidades de desarrollo.

8. Disponible en: https://www.socialfuturo.com/noticias-tecnologicas/microsoft-incorpora-la-tecnologia-de-chat-gpt-en-sus-productos-de-microsoft-365/#:~:text=La%20reconocida%20inte [fecha de consulta: 06/03/2024]

9. Disponible en: https://www.socialfuturo.com/noticias-tecnologicas/microsoft-incorpora-la-tecnologia-de-chat-gpt-en-sus-productos-de-microsoft-365/#:~ [fecha de consulta: 06/03/2024]

10. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=8103852 [fecha de consulta: 06/03/2024]

11. Disponible en: https://archive.is/xCyKU [fecha de consulta: 06/03/2024]

negligencia en la programación de los "algoritmos", ya que tratándose de posibles daños cometidos por una "inteligencia inmaterial" será arduo el debate respecto de a quien imputar los resultados dañosos y si será a título de dolo o culpa.[12]

La propia tecnológica Open AI admitió al presentar este desarrollo que GPT-4 tiene "habilidades de razonamiento más avanzadas" que Chat GPT, pero advirtió que aún podría ser propenso a compartir información errónea".[13]

En esta rápida y profunda evolución buscamos explicarnos, entonces no sólo por las causas que nos llevaron a este punto del desarrollo humano sino que también nos toca identificar las consecuencias de estos desarrollos, sin separar la descripción del fenómeno, de la cuestión valorativa respecto del uso ético de estas tecnologías con base en el respeto de los DD. HH.

Cabe señalar que el impacto de estos masivos y veloces cambios tecnológicos, según algunos autores, son mayores a las consecuencias que hace tres siglos produjo la revolución industrial, por lo que ya se habla de este fenómeno "como la cuarta revolución industrial".[14]

La IA, además de ampliar todos los límites del conocimiento humano y atraparnos con sus ilusiones, como afirma Harari, ya va produciendo consecuencias laborales con el consecuente desempleo producto del desplazamiento de puestos de trabajo. Ello nos enciende luces de alerta respecto del fenómeno, si no avanzamos en la forma señalada por los organismos internacionales hacia su regulación.

Ahora bien, muchos también se preguntan si es dable y conveniente para esta etapa de crecimiento de los desarrollos de IA pensar en trasladar la responsabilidad a los programadores y a las empresas

12. En torno a la imputación subjetiva, autoras como Lorena Varela, sostienen que algunos delitos más allá de un resultado típico y un nexo causal imputable, no requerirían *mens rea* como estado mental culpable en los términos del derecho anglosajón y cita varios ejemplos de "responsabilidad objetiva" en Derecho Penal.

13. Disponible en: https://www.socialfuturo.com/noticias-tecnologicas/microsoft-incorpora-la-tecnologia-de-chat-gpt-en-sus-productos-de-microsoft-365/#:~ [fecha de consulta: 06/03/2024]

14. Vega Iracelay, Jorge J, "Inteligencia artificial y derecho: principios y propuestas para una gobernanza eficaz", en *Revista Iberoamericana de Derecho Informático* (Segunda época), Federación Iberoamericana de Asociaciones de Derecho e Informática, 2018, pp. 13-14.

que lanzan estas herramientas al mercado si las mismas resultan defectuosas o insostenibles éticamente y debido a ello se corren serios riesgos de perjudicar derechos y bienes de terceros.

Esta pregunta se formula en un momento en que los países pretenden también afianzar cierto predominio tecnológico por sobre otros y mientras las empresas tecnológicas como Facebook anunciaron despidos masivos de personal.

Siempre las regulaciones marcharon muy por detrás de los cambios sociales y tecnológicos y mucho más cuando estos cambios resultan ser tan radicales.

En el marco supranacional en 2019 la OCDE[15] y los países socios, entre los que se encuentra la Argentina, se han suscripto formalmente una serie de recomendaciones generales sobre Inteligencia Artificial (IA), y han convenido en someterse a normas internacionales que velen por "el diseño de los sistemas de IA que los haga robustos, seguros, imparciales y fiables".

Resulta de nuestra preocupación que estas premisas no se perpetúen como una mera expresión de deseos, y que se ponga una intención nacional y global para desarrollar los sistemas de IA desde el principio con equidad (atendiendo a la no creación de instrumentos de análisis sesgados por prejuicios de raza, de género etcétera y con respeto de los principios de libertad e igualdad, lo que necesariamente implica la participación de todos los países y de todos los sectores involucrados: público, privado, académico incluyendo todas las perspectivas posibles).

Es por estos motivos que celebro que el Coloquio Preparatorio del XXI Congreso Internacional de AIDP sobre Inteligencia Artificial y Justicia Criminal, realizado en el ámbito de la Facultad de Derecho[16] parece haber dado un impulso nacional para una gran discusión, intercambio y regulación del fenómeno de la inteligencia artificial, no solo en el ámbito académico sino también en el sector de gobierno.

La falta de regulación en este caso sobre el fenómeno "algorítmico", o la "autorregulación" es un problema que venimos asistiendo des-

15. Disponible en: https://ia-latam.com/portfolio/principios-de-la-ocde-sobre-ia/ https://www.oecd.org/digital/artificial-intelligence/ [fecha de consulta: 06/03/2024]

16. Coloquio Preparatorio al XXI Congreso de AIDP sobre Inteligencia Artificial y Sistema de Justicia desarrollado en Buenos Aires entre los días 27 y 29 de marzo de 2023.

de la incorporación tardía de los delitos informáticos, al Código Penal Argentino en el año 2008,[17] casi treinta años después de la aparición de los primeros casos reveladores del fenómeno.

De modo que la aspiración es no llegar demasiado tarde con la regulación del fenómeno de la *Big Data* y de la Inteligencia Artificial.

Tal como puede leerse en la página oficial Argentina.gob.ar, la primera es el producto del desarrollo de las tecnologías de la información y la comunicación (TIC), que ha permitido la "generación, transmisión y almacenamiento de enormes cantidades de datos generados por diversas fuentes: sensores, dispositivos de red, páginas web, sistemas de correo electrónico, redes sociales, aplicaciones digitales" y toda la información almacenada en las cámaras de vigilancia en espacios privados y públicos más toda aquella información que puede brindar hoy el llamado Internet "de las cosas".

Esos conjuntos de datos son cada vez más grandes e inmanejables con los métodos y las herramientas tradicionales de procesamiento de datos; sin embargo, esto es posible dada la mayor capacidad de cálculo computacional de los procesadores en la nube y de las grandes empresas.

En conclusión, hay que decir que el manejo eficiente y justo de los datos se ha convertido en un gran desafío.[18]

La Unión Internacional de las Telecomunicaciones (UIT) define al *Big Data* como una práctica que permite la recopilación, el almacenamiento, la gestión, el análisis y la visualización, potencialmente en tiempo real, de amplios conjuntos de datos con características heterogéneas.

Por otra parte, con el desarrollo de las tecnologías digitales venimos utilizando en forma intercambiable términos como "algoritmos", los que de acuerdo a una definición posible, podrían definirse como "secuencias de pasos lógicos estructurados en programas" que permiten crear máquinas o instrumentos que presentan las mismas capacidades, o mejores, que las del ser humano; con lo cual podemos ver la inteligencia artificial como el resultado de la combinación de esos "algoritmos".[19]

17. Ley Nacional Nº 26388 sobre Delitos Informáticos, sancionada el 04/06/2008.

18. Disponible en: https://www.argentina.gob.ar/jefatura/innovacion-publica/telecomunicaciones-y-conectividad/grupo-de-trabajo-de-servicios-de-6 [fecha de consulta: 06/03/2024]

19. "Inteligencia Artificial: Peligro!", *Revista Viva la Ciudadanía*, Edición Nº 816, 11/03/2023.

En concreto, las tecnologías que permiten el procesamiento de grandes volúmenes de datos permiten resolver distintos retos en la menor unidad de tiempo, ampliando las capacidades humanas, ayudan a crear nuevas oportunidades de desarrollo y acceso al conocimiento en el sector público y el privado al tiempo que también son rápidamente aprovechadas como nuevas vías de oportunidad delictivas.

Los que venimos formando parte de las discusiones "ciber" estamos acostumbrados a debatir acerca de la responsabilidad por el delito computacional, informático o ciberdelito, cuando los ordenadores son utilizados como medio o como objeto de ataque a los datos, sistemas y bases de datos, y a una pluralidad de bienes jurídicos tutelados por el Derecho Penal.

Asistimos, por ejemplo, durante la pandemia de COVID-19 a la regulación de la actividad de ciber patrullaje[20] en redes sociales con el objeto de prevenir delitos de contenido (pornografía infantil y ciberterrorismo),[21] la puesta a disposición y venta de "falsos medicamentos contra el COVID y el ofrecimiento y distribución de medicamentes adulterados por la red", como así también al enorme desarrollo de sitios de suplantación de identidad con fines de "phishing" y fraude mediante la captación y utilización de nuestros datos personales, bancarios y crediticios.

Sin embargo, las herramientas utilizadas para la actividad preventiva en redes no han tenido un verdadero proceso de evaluación y control, por lo que estas permanecen aún opacas para la ciudadanía.

En el campo del derecho penal, resulta evidente que "los sistemas asistidos por IA y el análisis jurídico computacional, nos permite potenciar y enriquecer nuestras capacidades analíticas, suplementar nuestra capacidad de lectura, nos permiten analizar gran cantidad de datos "en crudo" y poder extraer variables, correlacionarlas y reconocer patrones, tendencias y movimientos dentro de la *Big Data* jurídica que constituyen las sentencias de nuestros tribunales y además ilumi-

20. Disponible en: https://observatoriolegislativocele.com/argentina-resolucion-144-2020-protocolo-de-ciberpatrullaje-en-fuentes-abiertas/ [fecha de consulta: 06/03/2024]

21. Parlamento Europeo, Reglamento Nº 784/2021.

nar el proceso de formación de las decisiones judiciales", tal como lo explica entre nosotros David Mielnik.[22]

Asistimos a las primeras investigaciones con producción y análisis de prueba informática, en las que se utilizan herramientas de extracción forense como el UFED de "Cellebrite",[23] que si bien no ha sido desarrollado en el país, la herramienta permite al investigador forense asociar programas para el análisis de imágenes y textos, correlacionando los datos de los dispositivos con otros datos provenientes de otras fuentes de libre acceso mediante una computadora (fuentes abiertas) y que pueden consistir además en fuentes no digitales.

¿Pero nos brindan las empresas desarrolladoras de estas herramientas de extracción y análisis la posibilidad de verificar y auditar los procesos internos que utiliza el algoritmo, nos permite saber qué datos de las personas se han correlacionado, para llegar a conclusiones fiables respecto de la prueba de la autoría y participación en ilícitos punibles? ¿Puede dicha prueba ser objeto de crítica por parte de las defensas? Esto es algo en el que el legislador procesal penal deberá pensar para futuras reformas procesales.

Los algoritmos no sólo mejoran la actividad de los abogados y de los jueces en el manejo de la "*Big Data* jurídica", sino que rápidamente estarán cambiando y en forma radical la actividad educativa en la que profesores y alumnos que ya están utilizando las herramientas de Microsoft Co: Chat bot GPT3 y 4 y los resultados de búsqueda de Bing que han incorporado un bot entrenado con técnicas de análisis del lenguaje natural, producto del *machine learning* o aprendizaje automático de las máquinas y del *deep learning* o "aprendizaje profundo".

Sin perjuicio de estos importantes desarrollos de Microsoft, por citar el ejemplo que ya todos conocemos y con el que de alguna manera ya hemos interactuado, al presentar la nueva herramienta, la jefa de IA responsable de Microsoft reconoció que el bot aún podía "alucinar" dando información falsa, afirmando sin embargo que la tecnología se había vuelto de todos modos "más confiable".

22. Mielnik, David, "Curso de análisis computacional del derecho penal argentino", Universidad Torcuato Di Tella.

23. Disponible en: https://cellebrite.com/es/ufed-ultimate-2/ [fecha de consulta: 06/03/2024]

En los días que siguieron al anuncio de Microsoft pudo observarse que Bing arrojo resultados con tremendos errores históricos.

Lo mismo se afirmó en relación con el desarrollo del Chat propio de Google (Bard) que en su primera demostración también arrojó errores fácticos y está al salir su nueva versión mejorada.[24]

Todos los errores son superados y se avanza en dichos desarrollos debido al enorme incentivo financiero que tienen las empresas para comercializar rápidamente las tecnologías digitales, sin que importe en ese afán, la regulación de todo lo vinculado a la preservación del anonimato, la seguridad y privacidad de nuestros datos, o las cuestiones éticas relativas a la no discriminación, y otros sesgos algorítmicos, la supervisión de errores etcétera; cuestiones que deberían ya estar incluidas en el diseño de estas herramientas, pero que todavía no sucede.[25] En igual sentido se pronunciaron recientemente técnicos de empresas informáticas e intelectuales que recientemente pidieron parar estos desarrollos por un tiempo.[26]

Resulta un lugar común que en nuestros países no se invierta mucho dinero cuando de responsabilidad o de seguridad de los datos se trata.

Segun Carly Kind, directora del Instituto Ada Lovelace, organización londinense enfocada en el uso responsable de la tecnología, la falta de regulación alienta a las empresas a "priorizar los intereses financieros y comerciales a expensas de la seguridad".[27]

Por un lado, la regulación parece llegar tarde[28] frente a la utilización de una cantidad de productos que ya son el resultado de los datos car-

24. Disponible en: https://www.wired.com/story/chatbots-got-big-and-their-ethical-red-flags-got-bigger [fecha de consulta: 07/03/2024]

25. Recomendación sobre la ética de la inteligencia artificial. Regulación de ética e IA de Europa y de la Unesco Publicado en 2022 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 7, place de Fontenoy, 75352 París 07 SP, Francia. UNESCO 2022.

26. Disponible en: https://www.elconfidencial.com/tecnologia/novaceno/2023-04-14/regulacion-inteligencia-artificial-peligro-chatgpt_3610493/ [fecha de consulta: 07/03/2024]

27. Disponible en: https://www.nuffieldfoundation.org/news/carly-kind-appointed-director-of-the-ada-lovelace-institute [fecha de consulta: 07/03/2024]

28. Proyecto de regulación de ética e IA de Europa y de la Unesco, *op cit.*, 19: allí se afirma que los sistemas asistidos por IA deben garantizar "la evaluación continua de la calidad de los datos de entrenamiento para los sistemas de IA, en particular la idoneidad de los procesos de recopilación y selección de datos, y que prevean medidas

gados en un algoritmo, que incluso ya han aprendido del contexto y son capaces de imitar tanto los textos, como las imágenes y la voz humanas.

Para todos esos avances se han utilizado terabytes de datos que son volcados en las redes por todos nosotros, usuarios de Internet, por lo que estas herramientas en manos de las plataformas digitales no reguladas o más bien autorreguladas, constituyen una amenaza para la integridad de nuestros datos, que ya vienen siendo utilizados para el desarrollo de estos grandes cambios tecnológicos y en estas nuevas aplicaciones.

Al igual que el delito informático, las técnicas algorítmicas tuvieron un pico de desarrollo y expansión durante la pandemia de SARS-CoV-2, los algoritmos utilizados por las empresas como Facebook, se dedicaban a moderar contenidos en redes sociales sin contar con el debido control humano, ya que en ese período empresas de tecnología funcionaban con un mínimo de personal y eran los algoritmos invisibles los encargados de acercarnos ciertos contenidos y censurar otros arbitrariamente, tal como se afirmó desde el "Observacom" en aquel momento:

Frente a la crisis sanitaria global por el Coronavirus, el uso de inteligencia artificial (IA) por parte de las redes sociales está mostrando, una vez más, sus limitaciones para la libertad de expresión: ya hay una gran cantidad de casos la remoción errónea de contenidos por supuestas violaciones a sus políticas.[29]

En virtud de estas alertas de organizaciones de la sociedad civil, las recomendaciones de organismos internacionales como la OCDE y el comienzo de regulación en Europa,[30] una reciente consulta con expertos realizada en Montevideo señaló la necesidad de generar un instrumento normativo que sirva a los gobiernos del mundo para diseñar políticas públicas respecto de este fenómeno y sobre todo abordar los principios éticos que atraviesan la disciplina.[31]

---

adecuadas de seguridad y protección de los datos, así como mecanismos de retroalimentación para aprender de los errores y compartir las mejores prácticas entre todos los actores de la IA".

29. Disponible en: https://www.observacom.org/censurabot/censura-algoritmica-covid19-expone-una-vez-mas-las-limitaciones-para-la-libertad-de-expresion-en-las-plataformas-de-internet/ [fecha de consulta: 07/03/2024]

30. Disponible en: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 [fecha de consulta: 07/03/2024]

31. Brito, Lidia, al inaugurar Foro de Montevideo en el primer encuentro virtual convocado por el gobierno de Argentina con apoyo de Unesco.

Las dificultades son evidentes: de un lado la inmensa cantidad de datos de todos nosotros que son utilizados para alimentar los algoritmos de IA, sin que a ciencia cierta se sepa cómo son adquiridos; y del otro, la dificultad para evaluar su impacto cuando las herramientas de IA ya fueron introducidas en el mercado, ello sumado a la imposibilidad de frenar u obstruir estos drásticos cambios y desarrollos tecnológicos que permiten un acelerado progreso humano y que a la brevedad se impone regular.

Vale la pena enfatizar en que el desarrollo de la IA no hubiera sido posible si no fuera porque desde hace años las plataformas en red (por ejemplo las del grupo Meta: Facebook, Instagram y WhatsApp) por citar sólo algunas de las que operan en nuestro país y crecen por el "efecto red", vienen trabajando hace largo tiempo con macro datos y algoritmos.

Estas plataformas tecnológicas como Facebook e Instagram localizadas en Sylicon Valley y otras como Twitter, crecieron por las expectativas de ganar dinero con publicidad y datos sobre todos nosotros, dado el volumen de usuarios que incorporan, al incluir, por ej. en Facebook a los "amigos" y "amigos de los amigos" en dicha plataforma y fueron aumentando su volumen a expensas de la inmensa cantidad de datos procesados de sus millones de usuarios. Este modelo de negocios permitió acumular ingentes ganancias[32] ya que cuantos más suscriptores tiene, más aumenta el interés por la plataforma.

Dado su crecimiento, las plataformas recurren a filtros de algoritmos en forma automatizada para, como dijimos moderar contenidos, y esa selección algorítmica ha demostrado "errores", que en muchos casos son directamente "censura". Por ejemplo, Facebook marcó como spam información verdadera sobre el coronavirus.

Desde Twitter distintos usuarios de diversas partes del mundo alertaron en la época de aislamiento social sobre la arbitraria remo-

32. El efecto red es un concepto financiero y se trata de una ventaja competitiva que se produce cuando la demanda y el valor de un proyecto, bien o servicio, aumenta en consonancia con el número de usuarios que tiene. Cuanto mayor sea esta base, mayor es la demanda y su valor añadido, lo que provoca la construcción de comunidades sólidas de usuarios. Este efecto resulta esencial para la expansión de un negocio, ya que cuantos más consumidores obtengan un determinado producto o servicio, la demanda de este también será mayor.
A modo de ejemplo, esto lo podemos ver muy claro en las plataformas de redes sociales: al principio la utilizan pocos suscriptores, conforme estos aumentan, el interés por la plataforma también lo hace.

ción de contenidos relacionados con el COVID-19 bajo la poco transparente causal de "violación de los estándares comunitarios", lo que sucede en las distintas redes que crecen a expensas de nuestros datos y que además se auto regulan.

La utilización de IA para estos procedimientos y la falta de transparencia en los criterios no son novedosas, pero en momentos excepcionales sus efectos se tornaron más visibles y sistemáticos, por lo que también se alertó desde la sociedad civil sobre los peligros de la "moderación privada".[33]

Los algoritmos son en definitiva datos que se categorizan, se personalizan, se correlacionan y se extraen patrones de conducta. A estos algoritmos se los entrena a través de operaciones repetitivas y las empresas tienen que ganar volumen de usuarios para hacer las operaciones cada vez con mayor cantidad de datos.[34]

Dada esta realidad, Harari[35] plantea que el mundo del siglo XXI podrá ser percibido como un "gran flujo de datos" y todas las decisiones serán parte de un procesamiento de datos en el que nos fusionaríamos, de hecho, agrega el historiador "hoy en día ya nos estamos convirtiendo en minúsculos chips dentro de un gigantesco sistema de procesamiento de datos que nadie entiende en realidad…".

Modelos algorítmicos deseables y no deseables

Vale la pena destacar que el uso de sistemas automatizados y el desarrollo de la IA basado en incesantes flujos de datos no hubiera sido posible sin el mencionado "efecto red" por el que empresas de propiedad del Grupo Meta (Facebook, IG y Whatsapp) por citar sólo algunas de las que operan en nuestro país y a nivel global, fueron creciendo a expensas de sus usuarios y ganaron ingentes sumas de dinero con la explotación de datos mediante la recolección, análisis y segmentación de contenidos (marketing/targeting de datos).

33. Disponible en: https://www.wired.com/story/chatbots-got-big-and-their-ethical-red-flags-got-bigger [fecha de consulta: 07/03/2024]

34. O'Neil, Cathy, *Weapons of Math Destruction: How Big Data increases inequality and Threatens Democracy*, New York, Crown Publishers, 2016, p. 272.

35. Harari, Yuval Noah, *21 Lecciones para el Siglo XXI*, Titivillus, 2019.

Este modelo de negocios, que como señale les permitió acumular ingentes ganancias, fue desplazando –en cuanto a ganancias– a otras tradicionales empresas petroleras, financieras o de energía.

En definitiva, la clave de todo este nuevo proceso de acumulación, son los algoritmos que utilizan las empresas tecnológicas, que son los "editores" de datos que se categorizan, se personalizan, se extraen variables, se comparan y se establecen patrones de conducta de todos nosotros para en definitiva "orientarnos" y "sugerirnos" consumos, lecturas, ideas, ideología, incluida la orientación de nuestro voto.

A estos algoritmos se los entrena a través de operaciones repetitivas y las empresas han debido ganar el volumen de datos de usuarios indicado, para hacer las operaciones basadas en la *Big Data*.

Empresas que, como Google, hacen constantes pruebas y controlan miles de variables, retroalimentan sus algoritmos con la opinión de millones de personas y hacen un seguimiento para pulir sus resultados, hace que lo que todo este procedimiento esté más cerca de una estadística que otro modelo menor, en el que los resultados del análisis dependen de los datos de 20 o 30 personas, los que por supuesto carecen de solidez estadística.

Los estadísticos usan grandes cifras precisamente para compensar las excepciones y las anomalías, y requieren una retroalimentación, algo que les indique cuando se están desviando y a su vez, se utilizan los errores para enseñar a sus modelos y hacerlos más inteligentes.

Las empresas se ocupan de corregir al algoritmo (por ej. el que nos recomienda libros) hasta hacer que funcione bien.

El hecho de que un algoritmo "funcione bien" se puede medir por la cantidad de *likes* en las redes. Un motor de búsqueda es un sistema estadístico. Estos sistemas trabajan así, para evitar hacer análisis defectuosos o perjudiciales.

En cambio, hay modelos a los que la científica de datos, Cathy O'Neil,[36] llama "armas de destrucción matemática" que son las que a menudo castigan a personas concretas que resultan ser la "excepción" del modelo.

36. O'Neil, Cathy, *Weapons of Math Destruction: How Big Data increases inequality and Threatens Democracy*, *op. cit.*

¿Y cómo lo hacen?

Pues definiendo su propia realidad y utilizándola para justificar sus resultados.

Este tipo de modelo se autoperpetúa y es altamente destructivo por no aprender nunca de sus errores. En vez de buscar la verdad, es el propio algoritmo el que la personifica.

Hoy el algoritmo no acerca información, sino que condiciona y orienta nuestras elecciones.

Un modelo no pernicioso trabaja con datos del pasado; no obstante, cada vez más se usan modelos que utilizan datos sustitutivos (esto es muy común tanto para tomar nuevos empleados, reducir personal como en los sistemas educativos para evaluar o incluso despedir docentes que, según el algoritmo, no llegan a un rendimiento adecuado), al decir de la citada científica.

Agrega Cathy O'Neil que cada vez más se utilizan informes de solvencia crediticia para llegar a conclusiones tales como qué determinado aspirante a un trabajo tiene mayor probabilidad de llegar temprano o de cumplir las normas, según pague o no sus facturas o si lo hace puntualmente. Con este razonamiento se estarían utilizando datos que son claramente sustitutivos y que por lo tanto se basan en una creencia, no en datos objetivos del pasado.

En efecto todos sabemos que la solvencia crediticia de una persona puede caer no necesariamente por cuestiones imputables a ella, dado que una persona puede comenzar a deber facturas de servicios o pagarlas fuera de término –tal como sucede en países como el nuestro con altos índices de pobreza, desigualdad y sujeto a crisis financieras y de desempleo cíclicas–.

Es interesante el aporte de la estadounidense que afirma que el empleo de datos o informes de solvencia para tomar empleados arrastrara a cientos de miles de personas a la pobreza, lo que es un círculo pernicioso porque ello empeorara aún más las calificaciones de solvencia de esas mismas personas y esto constituye lo que la científica llama "un bucle de retroalimentación pernicioso".[37]

37. Ídem.

Hay muchas premisas perniciosas que bajo las "matemáticas" se mantienen sin que se las verifique o cuestione (sesgo de automatización).

Estos modelos son como una caja negra y los resultados no pueden explicarse por sí mismos. Se trata de secretos corporativos. Resulta imposible discutir esos modelos o protestar en contra de ellos. Los programadores no abren los datos de sus sistemas y por lo tanto estos algoritmos perniciosos no pueden retocarse y sin embargo resultan inapelables por el referido "sesgo de automatización" ya que "si lo dicen las matemáticas estará bien".[38]

Entre los modelos predictivos y perniciosos aplicados en la Justicia Penal nos detendremos en los más sesgados o en los "algoritmos prejuiciosos".

Son los modelos "chapuceros" que utilizan datos sustitutivos, tanto para encontrar clientes, manipular personas desesperadas que buscan crédito o empleos y que en el caso de "la Justicia" sirven para "orientar" procesos de condenas a para privar a personas de su libertad o para determinar por ej., su capacidad de reincidir en el delito (modelo predictivo utilizado en la administración de justicia).

Esta oscuridad en la selección de los datos con los que se entrena al algoritmo precisamente es lo que la autora citada en último término llama "el lado oscuro del *Big data*".

En cambio, sigue explicando, que si a cada una de las variables de un modelo matemático se las conoce y se las puede cuantificar, se le adjudica un valor o un puntaje y se incluyen todas las relaciones medibles entre los distintos componentes, se analizan patrones, se efectúan comparaciones y se establece un complejo tapiz de probabilidades, ese modelo como la científica explica, puede funcionar dado que si los datos del pasado son "fiables o corroborables" su utilización sirve para mejorar los resultados de cualquier proceso para el que se los utilice.

De esta forma, con las combinaciones óptimas, los modelos predictivos ya no serían perniciosos porque los datos que toman son relevantes para los resultados que quieren predecir.

38. Lo que suele denominarse "sesgo de automatización" es la falta de escepticismo ante la información que nos proporcionan los algoritmos. Paradójicamente, aquí somos nosotros, y no las máquinas, quienes pecamos de actuar de forma automática. Y como todos los sesgos, tendemos a negarlos.

En general los modelos perniciosos reemplazan datos que no tienen de las personas por datos sustitutivos, generalmente porque no disponen de datos relativos a los comportamientos que más les interesan y así establecen correlaciones con datos como el código postal, sus patrones de lenguaje y de consumo, su solvencia crediticia y a estos datos se los correlaciona por ejemplo con su potencial para devolver un préstamo, o para cumplir con su trabajo, o para reincidir en el delito.

Estas correlaciones son discriminatorias[39] y algunas de ellas ilegales. Un modelo debe también poder ser comparado entre lo que predice y la realidad para identificar en que se han equivocado. Esos son los llamados por la analista matemática citada, modelos fiables.

Todos tenemos modelos que nos guían para la toma de decisiones y en general usamos datos conocidos del pasado y, aun así, hay una cantidad de incertidumbre respecto de la decisión.[40]

De todos modos, un modelo matemático es siempre una simplificación que no puede incluir la complejidad del mundo ni los matices de la comunicación humana, ni –agrego– conceptos tan abstractos como el de justicia. Ver por todos Cathy O'Neil, Blinik,[41] Zizek[42] por citar algunos autores provenientes de diferentes ciencias.

Así lo afirmó también recientemente el lingüista de la Universidad de Arizona Noam Chomsky,[43] quien refiriéndose a los chatbots y otros programas hermanos dice que necesariamente trasladan su responsabilidad a sus creadores por cuanto no saben equilibrar la creatividad con las restricciones éticas, las *fake news* o incluso las mentiras o datos falsos.

39. Disponible en: https://datagenero.medium.com/sesgos-en-los-algoritmos-d13898884cd9. [fecha de consulta: 06/03/2024]. En esta publicación se describen lo que son los sesgos algorítmicos y como se producen. Allí se afirma que "El sesgo algorítmico se puede presentar de diferentes maneras, como sesgo de género, sesgo racial, sesgo demográfico, sesgo económico, etcétera. Por lo general, el sesgo desfavorece a las minorías o a aquellos grupos que no están bien representados en los datos que se utilizan para entrenar modelos de aprendizaje automático".

40. O'Neil, Cathy, *Weapons of Math Destruction: How Big Data increases inequality and Threatens Democracy*, *op. cit.*

41. Disponible en: https://youtu.be/lqJODrHfAXE [fecha de consulta: 06/03/2024].

42. Disponible en: https://www.clarin.com/cultura/slavoj-zizek-inteligencia-artificial-peligro-tomar-chatbot-persona-personas-hablen-chatbots-_0_UGO4aDdnqs.html [fecha de consulta: 06/03/2024]

43. Disponible en: https://palabrapublica.uchile.cl/la-falsa-promesa-del-chatgpt/ [fecha de consulta: 06/03/2024]

O bien, continua el profesor, sobregeneran contenidos produciendo tanto verdades como falsedades, apoyando decisiones éticas y no éticas por igual o exhibiendo falta de compromiso con cualquier decisión e indiferencia a las consecuencias. Dado que estos sistemas ostentan amoralidad, falsa ciencia e incompetencia lingüística.[44]

La verdadera inteligencia también es capaz de pensar moralmente. Esto significa restringir la creatividad ilimitada de nuestras mentes con un conjunto de principios éticos que determina lo que debe y no debe ser (y, por supuesto, sometiendo esos principios mismos a la crítica creativa). Para ser útil, Chat GPT debe estar facultado para generar resultados novedosos; y para ser aceptable para la mayoría de sus usuarios, debe mantenerse alejado de contenido moralmente objetable (Chomsky y Zizek).[45]

Señalan los referidos profesores que, como no tienen capacidad de razonar a partir de principios morales, pueden arrojar resultados incluso ilegales si se lo carga de datos de entrenamiento ofensivo, misógino o racista, como en un pasado reciente sucedió.[46]

Estos sistemas son por diseño, limitados en lo que pueden aprender (o memorizar) son incapaces de distinguir lo posible de lo imposible (pueden aprender tanto que la tierra es chata como que es redonda). Chat GPT3 para el Profesor de Lingüística, exhibe plagio, apatía y obviedad.

Sin embargo, el chat bot GPT3 también al decir de Chomsky ha concertado enorme interés desde que Microsoft lo anunció, sin perjuicio de las luces rojas que despertaron los algoritmos de generación de texto, debido a sesgos y falsedades detectadas.

No obstante, y a pesar de estas anomalías, los gigantes tecnológicos se están apresurando a desarrollar todo tipo de herramientas de IA generativas de imágenes, de voz, de texto, a pesar de las objeciones éticas de los investigadores.[47]

---

44. Ídem.

45. Disponible en: https://www.clarin.com/cultura/slavoj-zizek-inteligencia-artificial-peligro-tomar-chatbot-persona-personas-hablen-chatbots-_0_UGO4aDdnqs.html [fecha de consulta: 06/03/2024]

46. Disponible en: ttps://www.wired.com/story/chatbots-got-big-and-their-ethical-red-flags-got-bigger/ [fecha de consulta: 06/03/2024]

47. Ídem.

Toda esta ingente masa de datos de la que se nutre la inteligencia artificial, es indudablemente más fácil de compilar computacionalmente que por un ser humano, en cada campo del conocimiento humano.

El gran tema a tener en cuenta para entrenar a un algoritmo en forma fiable es restringir dicho entrenamiento sólo a aquellos datos que puedan ser útiles y no sesgados.

En este ámbito las regulaciones son complejas, tal como lo fueron las de Internet en donde la normativa siempre llegó corriendo de atrás al fenómeno al que deseaban regular, y las empresas proveedoras de servicios de Internet trabajaron tanto para acelerar sus negocios como para desacelerar las regulaciones, so pretexto de que "la industria necesita pocos obstáculos, ya que Estados Unidos compite con China por el liderazgo tecnológico", tal como pudo leerse en *The New York Times*,[48] edición del 23 de marzo de 2023, donde se sostiene que Washington está tomando una postura de no intervención a este proceso de auge del desarrollo de la IA que se ha apoderado de Silicon Valley, con Microsoft, Google, Alphabet y Meta compitiendo entre sí para acelerar los desarrollos tecnológicos.

Así la generación de *chatbots*, los automóviles auto pilotados, las cortadoras de césped inalámbrica, por citar diversos artefactos culturales, vienen generando el debate comentado sobre el desarrollo tecnológico y los límites éticos, sin soslayar ciertos temores de que la robótica pueda eventualmente reemplazar a los humanos en los trabajos o incluso volverse conscientes.

La economía del *Big Data* ha dado ingentes riquezas a sus empresas, lo que las mantiene al tope de todos los ratings de ganancias, sin

48. Disponible en: https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html?unlocked_article_code=1TqqndCUgN8reUHIEiz0Z_ezwUCSq8Oh1O8CvEutu5yylo1Dno9gdFYm7CvpooLaoNNgkjoiuto1jPZEy4q3eVXYVuSy5H_wmlq2XpHFoiqpKKOgkbXdxA2r61WI3T6tfiJUXIBt_LDmK67wPvi2HmqeXze3kXCPuoD-lXV8mpUlc6SuSL6ObWWZGGM9bvOORRHX9rXnDOuDf_Fuv68dcCGTqgwoi9tHboN_lA-vhVb2Ct9ibNnzMPvwArEwEP4C-6nLxiaJ11duSYPOPfD6bLuSWhn8PN7DnJ7eAeSBR_WxwPMESwyFveCe2wRqRwmnwvjADQDAiDvSMitb5UolTETHhSrrhN2rUfGmh85V6mRCOPEgE9QZnA&smid=url-share [fecha de consulta: 06/03/2024]

perjuicio de lo cual recientemente hemos podido asistir a la caída del Banco de Silicon Valley (SVB).[49]

Lo cierto es que con un programa de computadora se pueden procesar miles de currículos, procesar solicitudes de préstamos en un par de segundos, datos de salud, etcétera y clasificarlos en listas bien ordenadas con los candidatos más prometedores ubicados en los primeros puestos de las búsquedas laborales, etcétera.

También es cierto que los programas informáticos que cruzan y parametrizan nuestros datos permiten sustituirnos en tareas que a los humanos nos lleva ingente cantidad de tiempo y que pueden ser rutinarias, monótonas y repetitivas.

¿Pero qué queda de la objetividad de la programación de un algoritmo? Y cómo hacer que ellos sean "objetivos y justos" tal como estos procedimientos fueron anunciados y que las decisiones superen o mejoren las humanas, en la medida en que no se utilicen herramientas automatizadas que muchas veces son "opiniones prejuiciosas programadas".

Parte de la solución es que los modelos matemáticos sean supervisados con visión ética y corregida cuando se producen desvíos y que no se programen los prejuicios, las equivocaciones y los sesgos humanos, pero bajo la falsa apariencia de que los modelos matemáticos son mucho más fiables que las decisiones humanas.[50]

La justicia predictiva basada en datos. Breve referencia al sistema COMPAS de evaluación de riesgo. Crítica al algoritmo de COMPAS

Hoy en día los Tribunales, bancos y otras instituciones están empleando sistemas automatizados de análisis de datos para tomar decisiones que afectan nuestras vidas y dichas decisiones son dejadas en manos de algoritmos que deciden las prioridades y que no solo nos informan, sino que fundamentalmente nos orientan.

49. Disponible en: https://www.infobae.com/economia/2023/03/14/las-razones-del-derrumbe-de-sillicon-valley-bank-un-experto-analiza-las-desacertadas-decisiones-del-banco/ [fecha de consulta: 06/03/2024]

50. O'Neil, Cathy, *Weapons of Math Destruction: How Big Data increases inequality and Threatens Democracy*, *op. cit.*

Esto es señalado por la revista *Mit Tecnologie Review* que nos alerta nuevamente sobre la discriminación, sesgos y censura algorítmica, esta vez refiriéndose al algoritmo de COMPAS,[51] que es una herramienta de evaluación de riesgos que se utiliza en el sistema de Justicia Penal de los EUA para medir la posibilidad de reincidencia, el que sin quererlo comenzó a discriminar a las personas por su raza y que fue detectado a través de un análisis de la fundación Pro Publica, que comenzó a comparar las evaluaciones de riesgo de COMPAS de más de 10.000 personas detenidas en un condado de Florida (EE. UU.) con la frecuencia con la que realmente volvían a reincidir.

Se explica en la publicación que al algoritmo de COMPAS se lo entrena con los datos históricos de los acusados para encontrar correlaciones entre factores como la edad, sus antecedentes en el sistema de justicia penal, y toma en cuenta también los arrestos policiales.

Luego usa esas correlaciones para predecir la probabilidad de que un acusado sea detenido por otro delito durante el período de espera del juicio.

Según su análisis, el algoritmo "predijo correctamente la reincidencia para los acusados blancos y negros con aproximadamente la misma tasa de éxito". Pero cuando el algoritmo fallaba, el error era distinto para negros y blancos.

En concreto, el error consistía en que "los negros son casi dos veces más propensos que los blancos a ser clasificados con un riesgo más alto sin llegar realmente a reincidir".

Ello ocurre porque estas herramientas utilizan los previos arrestos como un indicador de dicha probabilidad; pero, en este punto, hay grandes discrepancias entre negros y blancos, porque la policía tiene un historial desproporcionado de arrestos de minorías raciales y de manipulación de datos, se señala en la mencionada publicación.

Además, los nuevos arrestos de reincidentes ocurren a menudo por culpa de infracciones técnicas, como no comparecer ante el tribunal, más que por la reincidencia en la actividad delictiva.

51. Disponible en: https://www.technologyreview.es/s/7950/unamonos-pa [fecha de consulta: 06/03/2024]

Esto se llama puntuación de riesgo y el objetivo es actuar a modo de recomendación a los jueces para que los sujetos de alto riesgo sean encarcelados durante el proceso.

Lo que se demostró es que dentro de la zona de acusados de "alto riesgo" hubo algunos que no debieron ser realmente encarcelados y otro tanto ocurrió con los de "bajo riesgo".

Se podrá decir que este es un equilibrio con que el sistema de justicia penal siempre ha lidiado, así que la cosa no cambia cuando usamos un algoritmo (lo que sucede por las tasas de falsos positivos y de falsos negativos que arrojan).

En este caso sería ético seguir la máxima utilizada desde el siglo XVIII que reza: "Es mejor que 10 culpables sigan libres a que un inocente sufra".

No existe ningún algoritmo capaz de solucionar este problema. De hecho, ni siquiera es un problema algorítmico. Los jueces están tomando actualmente el mismo tipo de decisiones sesgadas o forzadas (cuando benefician o no, en función de raza, genero, incapacidad), y así lo han hecho a lo largo de la historia.

Pero hay algo que el algoritmo sí ha cambiado.

Aunque es posible que los jueces no siempre sean transparentes sobre cómo eligen entre diferentes nociones de lo justo, las personas pueden impugnar sus decisiones. Por el contrario, COMPAS, elaborado por la empresa privada Northpointe y según el MIT Technology Review, es un secreto comercial que no puede ser revisado ni interrogado públicamente.

Los acusados ya no pueden cuestionar sus resultados y las agencias gubernamentales han perdido la capacidad de analizar el proceso de la toma de decisiones. Ya no existe la rendición pública de cuentas, se señala en el informe de dicha institución.

La Ley de Responsabilidad Algorítmica propuesta en 2019[52] es un ejemplo de un buen primer paso, según el profesor de derecho de la Universidad de California y especialista en inteligencia artificial y derecho, Andrew Selbst, citado en el artículo al que nos referimos.

52. Disponible en: https://revistas.ucm.es/index.php/TEKN/article/view/79692 [fecha de consulta: 06/03/2024]

Este proyecto de ley, que quiere regular el sesgo en los sistemas automatizados de toma de decisiones, tiene dos características destacables que sirven como modelo para la futura legislación.

La primera, es que obligaría a las grandes empresas de tecnología a controlar sus sistemas de aprendizaje automático para detectar sesgos y discriminación en una "evaluación de impacto".

La segunda, es que no especifica la definición de lo justo (porque entiende que ello significa diferentes cosas en distintos contextos). Sin embargo, expresa Selbst que algunas "eficiencias" en la construcción del algoritmo ameritarían encontrar tales definiciones en cada caso.

Al fin y al cabo, los sistemas predictivos se basan en estadísticas generalizadas, no en la situación individual de alguien.

Esto siempre ha sido ajeno al Derecho Penal en el que se evalúa la responsabilidad personal por el hecho; no obstante, podrían ser asistentes útiles para lograr decisiones más sabias y justas que las que los seres humanos hacen por sí solos, pero siempre y cuando se tengan en cuenta las referidas pautas relativas a la evaluación de impacto y se eliminen los cuestionamientos relativas a la falta de control de los desvíos y se estudie la incorporación del concepto de "lo justo" adaptado al caso concreto. Saldemos antiguas discusiones antes de regular lo nuevo.

Lo que nos convoca como juristas es preguntarnos tanto por el entrenamiento algorítmico, sus sesgos y errores o desvíos para la evaluación de riesgos como por las responsabilidades por los resultados erróneos o dañosos producidos por una AI y el título de su imputación.

Esto se respondería, en parte, saldando el debate sobre la responsabilidad civil, administrativa o penal de los ISP, que debido a la cantidad de información que manejan, junto a su capacidad de almacenamiento y de cómputo, son las que tienen las condiciones para desarrollar estas herramientas, su programación y su eficiente y no delictiva introducción al mercado.

Obviamente les corresponde una responsabilidad civil e incluso penal por cuanto son los ISP los que explotan económicamente nuestros datos.

Recientemente pudo leerse en *The New York Times*:

> Que cuando los legisladores destacan los avances de la tecnología, pocos están tomando medidas al respecto. No se ha propuesto ningún proyecto de ley para proteger a las personas o impedir el desarrollo de los

aspectos potencialmente peligrosos de la IA. Y la legislación introducida en los últimos años para frenar las aplicaciones de IA como el reconocimiento facial se ha marchitado en el Congreso.[53]

Es algo sabido que falta el debate a fondo respecto de por lo menos tres puntos centrales:

a. Como se distribuyen las responsabilidades en el mundo digital.
b. Cuál debe ser la sanción por las consecuencias de la utilización no transparente de nuestros datos.
c. De quien es la propiedad de los datos personales, los almacenados por las empresas y de quien son los contenidos puestos en las redes por los usuarios.

Somos los usuarios dependientes de Internet, los que suministramos los datos con los que las empresas alimentan las herramientas de IA y no tenemos injerencia alguna en la forma en que son procesados y menos que menos en los tipos de resultados que arrojan de nosotros los consumidores y usuarios de la tecnología.

Baste recordar que uno de los efectos perniciosos de la *Big data* que fueron debatidos y no solucionados, a raíz de lo testimoniado por ex directivos de las principales empresas tecnológicas, que sacaron a la luz los mecanismos depredadores de las grandes plataformas y redes sociales para lograr por ejemplo, la dependencia de niños, niñas y adolescentes respecto de sus dispositivos móviles y tenerlos permanentemente siguiendo la actividad en las redes.

Fue el presidente de los EUA, Biden, quien pidió al Congreso la urgente regulación de este grave problema.[54]

---

53. Disponible en: https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html?auth=login-google1tap&login=google1tap&smid=url-share&unlocked_article_code=1TqqndCUgN8reUHIEiz0Z_ezwUCSq8Oh1O8CvEutu5yylo1Dno9gdFYm7CvpooLaoNNgkjoiuto1jPZEy4q3eVXYVuSy5H_wmlq2XpHF0iqpKKOgkbXdxA2r61WI3T6tfiJUXIBt_LDmK67wPvi2HmqeXze3kXCPuoD-lXV8mpUlc6SuSL6ObWWZGGM9bvOORRHX9rXnDOuDf_Fuv68dcCGTqgw0i9tHboN_lAvhVb2Ct9ibNnzMPvwArEwEP4C-6nLxiaJ11duSYPOPfD6bLuSWhn8PN7DnJ7eAeSBR_WxwPMESwyFveCe2wRqRwmnwvjADQDAiDvSMitb5UolTETHhSrrhN2rUfGmh85V6mRCOPEgE9QZnA [fecha de consulta: 06/03/2024]

54. Disponible en: https://tekiosmag.com/2023/02/28/la-urgencia-de-joe-biden-por-regular-como-los-ninos-usan-la-tecnologia/

También resulta un testimonio alarmante de la carencia de regulación y control, del uso indebido, manipulación y cesión no autorizada de nuestros datos, el mal uso de Facebook (hoy Meta Platforms) que cedió sin autorización datos de los votantes estadounidenses para favorecer a candidatos políticos de un partido, manipulando los resultados de diferentes elecciones.[55]

Sin embargo, es poco lo que se ha avanzado en el control y sanción penal a las empresas, capaces de ceder sin autorización nuestros datos y atraer a niños y adolescentes en las redes un promedio 9 horas diarias[56] mediante la estrategia de enviar permanentes notificaciones de noticias de sus "amigos", para tenerlos en vilo y dependientes para así poder mandar publicidades y medir contenidos para niños y adolescentes.[57]

Estas conductas perniciosas[58] se suman a los ya referidos problemas de sesgos, omisiones e incorrecciones de los algoritmos que tienden a ahondar y profundizar diferencias sociales, raciales, religiosas, de género, políticas y económicas, entre otras, estableciendo profundas fragmentaciones sociales, sin que hasta el momento estos efectos y la responsabilidad por la programación defectuosa de los algoritmos hayan sido regulados.

Obviamente se suman a las víctimas tradicionales, los niños y niñas, las personas no nativas digitales y los usuarios en general, que como consumidores de Internet, nos hemos visto obligados a migrar nuestras interacciones y operaciones a las diversas plataformas digitales para desarrollar la totalidad de nuestras interacciones personales, trámites y transacciones.

Todo este avance tecnológico generó estas nuevas víctimas que por falta de capacitación e información caen en manos de terceros inescrupulosos que operan dentro de este entorno.

55. Disponible en: https://elpais.com/retina/2020/10/15/tendencias/1602775507_386132.html [fecha de consulta: 06/03/2024]

56. Disponible en: https://cnnespanol.cnn.com/2015/11/03/los-adolescentes-pasan-9-horas-al-dia-usando-los-medios-segun-informe/ [fecha de consulta: 06/03/2024]

57. Disponible en: https://www.unicef.org/uruguay/redes-sociales-y-adolescentes-lo-que-tenes-que-saber [fecha de consulta: 06/03/2024]

58. Disponible en: https://www.pagina12.com.ar/372879-frances-haugen-denuncio-los-efectos-nocivos-de-facebook-ante [fecha de consulta: 06/03/2024]

Además de los ya tradicionales delitos informáticos, que tienen a los sistemas informáticos como objeto o como medio de ataque, venimos asistiendo al desarrollo de la inteligencia artificial generativa de aplicaciones que hace tiempo se vienen entrenando con nuestras voces, textos e imágenes, por lo que prontamente deberán también regularse las consecuencias de su uso delictivo para el fraude y el plagio, y para posibles secuestros extorsivos y nuevas estafas cometidas mediante el uso de estas refinadas y convincentes herramientas que hoy nos envuelven.

Consideraciones preliminares

La tecnología basada en modelos estadísticos que muchas veces son opacos (modelos preventivos policiales como los de reconocimiento facial, hoy suspendido en la Ciudad Autónoma de Buenos Aires)[59] y de justicia predictiva –como los de medición de reincidencia comentados– no han sido regulados, a pesar de las grandes críticas formuladas por organizaciones de derechos humanos, en la medida que comprometen la intimidad, privacidad y seguridad de nuestros datos y que incluso cercenan nuestra libertad ambulatoria, la garantía de igualdad y la de prohibición de autoincriminación.

Las críticas pasan sin duda, por el uso masivo e indiscriminado de nuestros datos para entrenar a los algoritmos, y por los errores y sesgos que arroja su utilización en la programación de datos que son verdaderos "prejuicios codificados" lo que ocurre –en principio– por la falta de la debida supervisión y medición de impacto, que es la operación que precisamente permite corregir dichos desvíos.

Es dable advertir que este debate debe ser dado en el ámbito local pero también en el marco hemisférico y global porque las nuevas aplicaciones son utilizadas y despliegan sus contenidos en forma global, dada precisamente las características transnacionales de estos desarrollos.

59. Disponible en: https://www.cels.org.ar/web/2022/09/una-jueza-declaro-inconstitucional-el-uso-del-sistema-de-reconocimiento-facial-en-caba/ [fecha de consulta: 06/03/2024]

Conclusiones

En el afán de abarcar este fenómeno de *Big data*, Argentina cuenta con desarrollos en el sistema de justicia, alguno de los cuales se encuentran operativos y otros no.

En este sentido, y sin pretender abarcar a todos ellos, se destacan el programa de Análisis Computacional desarrollado por David Mielnik consistente en la búsqueda y análisis de las sentencias de las cuatro salas en la Cámara de Casación Penal Federal, el que sin perjuicio de no encontrarse operativo por el momento, propone un análisis computacional de dichas sentencias, automatizando en primer lugar la operación de descarga de fallos, que luego son almacenados y posteriormente la herramienta permite extraer texto e información básica de los mismos, lo que permitirá tanto a los jueces como a los letrados tomar decisiones basadas en las tendencias de la Cámara.

De esta manera –explica el autor– "se produjo una base de datos conformada por 44027 decisiones jurisdiccionales dictadas por las cuatro salas de la Cámara Federal de Casación Penal y la Sala de feria".

Señala Mielnik que el análisis se produce a una escala nunca vista, tanto por la posibilidad que nos brindan estas herramientas de ampliar nuestro campo de lectura reduciendo a días lo que llevaría semanas por lo que también permite conocer más profundamente cómo se van formando las decisiones judiciales: lo que el autor llama "iluminar el proceso de formación de las decisiones judiciales".[60]

Otra herramienta desarrollada por un físico de Conicet permitió reconstruir y desentrañar con imágenes de diversas fuentes –como fotografías de cronistas y videos de cámaras de seguridad de la Policía Federal, señales de televisión y otras–, las circunstancias en las que se produjeron cinco asesinatos y resultaron heridas varias personas durante la protesta social del 20 de diciembre de 2001 que tuvo lugar en la Ciudad de Buenos Aires.

Con el programa al que llamó "panóptico audiovisual" se pudieron analizar órdenes transmitidas por radio para así reconstruir lo ocurrido en esos días de diciembre de 2001 y poder identificar responsables

60. Mielnik, David, *Análisis computacional del Derecho Penal Argentino*, Tesis Magíster en Derecho Penal de la UTDT.

tanto políticos como materiales de la jornada. El programa permitió ordenar 300 horas de video y 500 fotografías, lo que no solo ayudó a querellantes, fiscales y defensas a revisar y documentar desde su punto de vista qué es lo que ese día ocurrió, sino también, que generó otro grupo de causas judiciales a partir de dicho material.

El UFED de la empresa Cellebrite[61] es una herramienta de forensia muy utilizada en los tribunales de nuestro país, y que actualmente permite distintos tipos de extracción de contenido de los dispositivos sometidos a pericia forense.

Se utiliza para la extracción lógica de los datos de un dispositivo móvil (SMS, registro de logs de llamadas, imágenes, agenda, videos, audios, ciertos datos de aplicaciones y más).

Extrae datos embebidos en la memoria de un dispositivo móvil, base de datos no visibles, passwords, y desbloquea y muestra contraseñas de una fuente como un dispositivo móvil. Puede clonar la tarjeta sim, es decir copia una ID de SIM de una tarjeta SIM a otra tarjeta SIM o a Cellebrite. Además, permite realizar la extracción física de una imagen bit a bit de la memoria flash de un dispositivo incluyendo el espacio no asignado, que es el área de la memoria flash que ya no es rastreada por el sistema de archivos, y que puede contener imágenes, videos, archivos etcétera. Asimismo, utilizando métodos avanzados de *carving*, –un proceso empleado para extraer una colección de datos de un conjunto más grande de datos–, se puede además agregar notas del investigador que lo utiliza.

El problema está cuando no queda claro si el comportamiento declarado de las herramientas resultara ser igual a los objetivos perseguidos y a lo que sucede realmente con el aprendizaje automatizado de las maquinas. Lo ético de estos desarrollos resulta del hecho de poder conocer siempre cómo son tomados los datos y para que la libertad probatoria en un proceso sea tal, es importante que no sean admitidas pericias que no pueden ser controladas por las partes.

A nivel administrativo existe un esfuerzo regulatorio por parte del BCRA dentro del sistema financiero y bancario argentino, y que mediante la reciente comunicación "A" 7724, referida a las herramientas de inteligencia artificial y su uso en el sistema financiero y bancario

61. Disponible en: https://cellebrite.com/es/ufed-ultimate-2/ [fecha de consulta: 06/03/2024]

obliga a las entidades a identificar y documentar el objetivo del uso, por sí o por terceros, de *software* que utilice algoritmos de inteligencia artificial o aprendizaje automático en sus proyectos o procesos.

Además, obliga a establecer roles y responsabilidades para la definición del contexto en que operan los sistemas de inteligencia artificial o de aprendizaje automático, a la identificación de los modelos, algoritmos y de los conjuntos de datos utilizados, objetivo del uso y la definición de Inteligencia artificial o aprendizaje automático y la definición de métricas y umbrales precisos para evaluar la confiabilidad de las soluciones implementadas. Asimismo, la comunicación prevé que deberán hacerse los análisis de riesgos correspondientes previendo las consideraciones mínimas que deberán medir; como por ejemplo los datos utilizados para el entrenamiento, su volumen, complejidad y obsolescencia, la privacidad y la afectación a los usuarios en su calidad de consumidores y otros establecidos.

Adicionalmente, prevé la obligación de implementar procesos que promuevan la confiabilidad en el uso de este tipo de algoritmos e incluyan al menos:

- Medidas para evitar la existencia de sesgos o discriminación contra grupos o segmentos de clientes o usuarios de los productos y/o servicios financieros.
- Documentación respecto de la transparencia, la explicabilidad de los modelos utilizados y la interpretabilidad de los resultados.
- La ejecución de revisiones periódicas de los resultados respecto de la tolerancia al riesgo definida.
- La comunicación al cliente cuando utilice servicios soportados por este tipo de tecnología.

Toda esta regulación de la IA resulta compatible con las disposiciones de la Agencia Federal de Datos de nuestro país y con las disposiciones del GDPR[62] europeo de 2018 respecto de la protección de datos personales. Lo que falta:

a. La regulación general del fenómeno siguiendo el aforismo latino: *nulla coactio sine lege* que muy especialmente contemple

62. Disponible en: https://nic.ar/es/enterate/novedades/entra-en-vigencia-nuevo-gdpr [fecha de consulta: 06/03/2024]

–entre otros problemas– los de privacidad/intimidad/igualdad y no discriminación como así también reglamentar la utilización transfronteriza de nuestros datos.

b. Como el fenómeno algorítmico es un fenómeno de *Big data*, por lo tanto imposible de abarcar con nuestras capacidades innatas, y dado que el Derecho es un fenómeno de *Big Data*, tenemos que incorporar los fantásticos beneficios del análisis jurídico computacional tanto para la resolución de casos simples como para la selección y análisis de fallos de los Tribunales y pensar como incorporamos a ellos el concepto de justicia –en el caso concreto– sobre cuya relatividad, según el contexto, resulta tan difícil de asir e incorporar al puntaje o valuación en un algoritmo.

Los algoritmos nos permiten hacer las mismas operaciones critico-jurídicas a escala donde quienes tienen la capacidad de acercar las fuentes de información son las m**á**quinas (Mielnik).

Introducir el análisis computacional del Derecho facilitada por los algoritmos no modifica lo que los abogados hacen actualmente, sino que altera la manera en que lo hacen e introducen a los juristas en nuevas y ricas fuentes de información.

Si bien no existe legislación específica de responsabilidad de las empresas que producen sistemas basados en inteligencia artificial, la doctrina nacional y los fallos a través de distintas pautas que van dando respecto de la responsabilidad, afirman que rigen las disposiciones relativas al consumidor[63] y a la responsabilidad por el producto.

La utilización de herramientas de IA en el proceso penal debe respetar la garantía de juicio justo, derecho al recurso y principio de inocencia, sobre todo cuando se trabaja con algoritmos que utilizan datos sustitutivos.

La libertad probatoria implica la utilización de pericias asistidas por IA de las que nos podamos defender.

Diversas agencias como la Red Iberoamericana de Protección de Datos elaboró dos documentos orientadores para el adecuado uso de

63. Disponible en: http://www.saij.gob.ar/corte-suprema-justicia-nacion-federal-ciudad-autonoma-buenos-aires-claps-enrique-martin-mercado-libre-sa-danos-perjuicios-fa13000193-2013-11-19/123456789-391-0003-1ots-eupmocsollaf [fecha de consulta: 06/03/2024]

datos personales en el diseño e implementación de inteligencia artificial, así como la Disposición 60/2016 de la Agencia de Acceso a la Información para la transferencia internacional de datos que, entre otras cosas, establece que de existir cualquier controversia sobre la información, la misma debe resolverse por ley argentina y por un juez argentino. A eso hay que sumarle las exigencias del GDPR[64] europeo de 2018.[65]

Esta disposición de la Agencia citada es de suma importancia, por cuanto también dispone, entre los Controles y notificaciones, que las organizaciones

> ... deberán proteger los datos personales con las medidas de seguridad apropiadas, notificar a las autoridades las filtraciones de datos personales, obtener los consentimientos apropiados para el procesamiento de los datos y mantener registros detallados de su procesamiento.

Para concluir cabe señalar que tal como fuera anunciado por Yuval Noah Harari en su libro *21 Lecciones para el siglo XXI*,[66] la libertad y la igualdad son los grandes temas del siglo y Harari lo expresa así:

> ... la fusión de la infotecnología y la biotecnología podrán hacer que muy pronto miles de millones de humanos queden fuera del mercado de trabajo y socavar tanto la libertad como la igualdad [...]
> Los algoritmos de macrodatos pueden crear dictaduras digitales en las que todo el poder esté concentrado en las manos de una élite minúscula al tiempo que la mayor parte de la gente padezca no ya explotación, sino algo muchísimo peor, irrelevancia.

64. El 25 de mayo del año 2018 entró en vigencia el Reglamento General de Protección de Datos (GDPR por su sigla en inglés). Se trata de una normativa de privacidad dictada por la Unión Europea que establece mayores derechos para los titulares de datos y mayores obligaciones para las personas, empresas, gobiernos y organizaciones que ofrezcan bienes y servicios o que recopilen y analicen datos vinculados a los residentes de la UE. Entre las novedades que incorpora el Reglamento, están las siguientes: "Privacidad personal: En relación a los datos personales, el GDPR establece que las personas tienen derecho a acceder a ellos, corregir errores que puedan tener, borrarlos, objetar su procesamiento y exportarlos".

65. Disponible en: https://es.unesco.org/news/inteligencia-artificial-america-latina-debate-normativa-abordaje-eticodisciplina [fecha de consulta: 06/03/2024]

66. Harari, Yuval Noah, *21 Lecciones para el Siglo XXI*, *op. cit.*

La utilización de un agente encubierto con Inteligencia Artificial a la luz de las garantías constitucionales

Brenda Flesler*

Introducción

La irrupción de las tecnologías modernas ha brindado un espacio con mayor facilidad para la comisión de delitos, entre ellos, el acceso, la divulgación y la distribución de representaciones de un menor dedicado a actividades sexuales explícitas e incluso a la posibilidad de que los pedófilos contacten a menores de edad desde un completo anonimato. Hace no mucho tiempo atrás, era necesario para realizar estas conductas introducirse en redes complejas de manera presencial, pero, con el advenimiento de internet, esta situación cambió sustancialmente. Este medio se convirtió en el ideal para la operación de redes complejas y cerradas en la que la producción, financiamiento, comercialización y publicación de material con contenido sexual infantil o el contacto con menores de edad con fines sexuales se hace de un modo en el que las identidades pueden mantenerse en total anonimato generando un riesgo mucho menor para los autores. Este escenario impuso la necesidad de todos los Estados de detectar maniobras ilícitas y contar con los medios y respuestas eficaces para combatirlas.

Es en este contexto, que en distintas legislaciones se incorpora, no sin ciertos reparos, la figura del agente encubierto. Hay autores como Maier o Sancinetti[1] que sostuvieron que se trata de un mecanismo por el cual el Estado se autoriza a cometer delitos y luego se perdona a sí

* Abogada con orientación en Derecho Penal por la Facultad de Derecho, Universidad de Buenos Aires. Magíster en Derecho Penal por la Universidad Austral (título en trámite). Magíster en Derecho Penal por la Universidad de Salamanca. Docente de derecho penal y Procesal Penal, Facultad de Derecho, Universidad de Buenos Aires. Empleada del Ministerio Público de la Defensa.

1. Rendo, Ángel Daniel, *Agente encubierto*, Tucumán, Editorial Albremática, 2010.

mismo, introduciendo de esta forma criminales autorizados a delinquir por resolución fundada. Sin embargo, de la vereda de enfrente se ha ubicado la Corte Suprema de Justicia Argentina (en adelante CSJN), en el marco del fallo "Fiscal c/ Fernández Víctor". Allí en los considerandos 10, 11, 12 y 13 desarrolló un minucioso análisis de esta técnica de investigación y sostuvo que su utilización para la averiguación de delitos no es por sí mismo contrario a garantías constitucionales. Y que una cuidadosa comprensión de nuestra vida social común comprueba que hay ciertos delitos graves que se preparan y ejecutan en una esfera de intimidad e imponen reconocer que solo son susceptibles de ser descubiertos mediante la utilización de infiltraciones en esos círculos.[2]

En este escenario ya poco pacifico, y frente a una realidad social en la que internet ofrece un estado de completo anonimato y la dificultad que eso conlleva de poder prevenir, advertir y evitar ciertos delitos cometidos en esos ámbitos antes de que estos sean consumados y resulten impunes, se alza la necesidad de adecuar las técnicas de investigación y prevención de los Estados a este nuevo tipo de delitos. La figura del agente encubierto no queda exenta a esta necesidad de modernización. Su clásica regulación y utilización como aquel funcionario policial infiltrado de manera presencial en cierto grupo de crimen organizado muchas veces resulta insuficiente en un escenario en el que los delitos no se cometen bajo esa modalidad, sino de forma virtual.

En este contexto irrumpe la inteligencia artificial que hace tiempo se encuentra aplicándose paulatinamente a los procesos y exige, necesariamente, una revisión de la regulación legal del instituto para que resulte aplicable.

El agente encubierto en la legislación argentina. Requisitos legales y principios constitucionales que la rigen

Con la sanción de la Ley Nº 27319, el legislador ha regulado la intervención del agente encubierto para delitos denominados "complejos"

2. CSJN, Fallos: 313-2:1305, "Fiscal c/ Fernández, Víctor Hugo s/ av. infracción Ley N° 2077", 11/12/1990.

especificados en la norma: narcotráfico (Ley Nº 23737); contrabando (Sección XII, título I del Código Aduanero); terrorismo (Art. 41 quinquies del Código Penal); corrupción de menores y su financiación, promoción y explotación de la prostitución; producción, financiación, ofrecimiento, comercialización, publicación, facilitación, divulgación, distribución toda representación de un menor dedicado a actividades sexuales explícitas con fines sexuales, y organizar espectáculos en vivo de representaciones sexuales explícitas en que participaren dichos menores (arts. 125, 125 bis, 126, 127 y 128, respectivamente del Código Penal); secuestros (arts. 142 bis, 142 ter y 170 del Código Penal); trata de personas (arts. 145 bis y ter del Código Penal); delitos cometidos por asociaciones ilícitas (arts. 210 y 210 bis del Código Penal); y delitos contra el orden económico financiero (Libro Segundo, Título XII del Código Penal).

Dentro de esta ley se establecen los requisitos que debe reunir una persona para poder ser designada como agente encubierto. Esto es, la pertenencia a las fuerzas de seguridad, estar en actividad y la voluntad de ser designado para tal laboral, estando prohibida la obligación a un funcionario de realizarla sin su consentimiento.

En cuanto al modo de llevar a cabo esta operación, el art. 4 de la Ley determina que una vez dispuesta la actuación por el juez, de oficio o a pedido del Ministerio Público Fiscal, su designación y la instrumentación necesaria para su protección estarán a cargo del Ministerio de Seguridad de la Nación, con control judicial. Dicho Ministerio es el organismo que seleccionará, capacitará, designará y protegerá al personal destinado a cumplir tales funciones. Para ello, cada fuerza policial deberá confeccionar una lista de carácter confidencial de funcionarios idóneos que se hayan postulado, aprobado la capacitación y se encuentren en condiciones de ser designados como agentes encubiertos.

También la norma establece las facultades de la autoridad judicial en el marco de esta técnica de investigación y se impone al juez la fundamentación de la medida y la obligación de regirse por los principios de necesidad, razonabilidad y proporcionalidad.

El presupuesto derivado de esta exigencia es que se encuentre abierta una investigación judicial relacionada con la comisión de alguno de los delitos complejos enunciados en la ley y que en el marco de ese caso concreto se realice un juicio de proporcionalidad del cual se concluya la imprescindibilidad de la medida.

Con la exigencia de este juicio y la concurrencia de los principios de necesidad, razonabilidad y proporcionalidad la ley prescribe que el empleo del agente encubierto es de carácter excepcional. Esto se debe a la alteración de principios constitucionales básicos y la intromisión a determinados derechos fundamentales que su actuación conlleva, razones determinantes para que su empleo quede sometido al cumplimiento de estrictos requisitos legales en respeto a las garantías procesales vigentes en un Estado de derecho.

Las regulaciones del agente encubierto en la legislación argentina y española. Diferencias y puntos en común

Como fue desarrollado en el punto anterior, la figura del agente encubierto fue incorporada a la legislación argentina a través de la Ley Nº 27319 que en su artículo 3 lo define como

> … aquel funcionario de las fuerzas de seguridad autorizado, altamente calificado, que presta su consentimiento y ocultando su identidad, se infiltra o introduce en las organizaciones criminales o asociaciones delictivas, con el fin de identificar o detener a los autores, partícipes o encubridores, de impedir la consumación de un delito, o para reunir información y elementos de prueba necesarios para la investigación, con autorización judicial.

Por su parte, en la legislación española la figura se encuentra regulada en el artículo 282 bis de la Ley de Enjuiciamiento Criminal. Con un criterio similar a la ley argentina, se regulan allí las cuestiones relativas a los requisitos de su actuación y el modo de llevarla a cabo.

Una y otra legislación reconoce como primer punto la necesidad de que la medida sea dispuesta por autorización judicial –en el caso de Argentina– o también por el fiscal dando inmediato aviso al juez –en el caso de España–. Esto va de suyo si tenemos en cuenta que la medida implica la intromisión en una esfera de intimidad de quien podría resultar autor de un delito, e incluso, en otras garantías constitucionales como la de no auto incriminarse o la inviolabilidad del domicilio. De allí, que la base imprescindible en este tipo de instituto resulte la auto-

rización de un juez; único modo legal de restringir aquellos derechos protegidos constitucionalmente.

Además, como derivación del propio principio republicano de gobierno, la orden debe encontrarse suficientemente fundada y superar un juicio de proporcionalidad.[3] Esto es: analizar si esta medida de carácter extraordinario resulta, en el caso concreto, necesario, proporcional, motivado y eficaz.

Otro punto en común que encuentran ambas regulaciones gira en torno a la eximición de responsabilidad del funcionario policial que actúe como agente por aquellas acciones que sean consecuencias necesarias de su actuación.[4]

Sin embargo, en este caso con ciertas diferencias en la técnica legislativa, la norma argentina establece un límite a la eximición de punibilidad en el caso en que la actividad ilícita por parte del agente implique un peligro cierto a la vida o integridad psíquica o física de una persona o a la imposición de un grave sufrimiento físico y moral. Es decir que esta cláusula de no punibilidad opera siempre y cuando la acción ilícita llevada a cabo por el agente en razón de su actuación no implique una afectación jurídica a la vida o integridad física de otra persona.

Por su parte, en la legislación española la regulación es más genérica. Se exime de responsabilidad al agente por aquellas actuaciones que sean consecuencia necesaria se su actuación, siempre que guarden la debida proporcionalidad con la finalidad de esta y no constituyan una provocación al delito. Como contracara, el límite a esa eximición de responsabilidad estará dado en aquellos casos en que las acciones ilícitas llevadas a cabo no resulten proporcionales con la finalidad de su actuación, o bien, sean consecuencia de una provocación a un ilícito.

Por último, ambas legislaciones establecen la posibilidad de que el agente declare en la etapa de juicio. Sin embargo, se diferencian entre sí en tanto Argentina reconoce expresamente la excepcionalidad de esa declaración, que solo será llevada a cabo cuando resulte

3. Lafont Nicuesa, Luís, *El agente policial encubierto*, Bilbao, Tirant Lo Blanch, 2022, p. 153.

4. Sobre el carácter extraordinario de esta medida ha sostenido, por ejemplo, la STS en el caso N° 104/2011 del 1ero de marzo que debe autorizarse la intervención de un agente encubierto "cuando han fracasado otros métodos de la investigación o estos sean manifiestamente insuficientes para su descubrimiento". Disponible en: https://vlex.es/vid/267170082

absolutamente imprescindible. Establece además que, en aquellos casos en que esa medida pueda poner en peligro la vida o integridad física del funcionario policial, deberán llevarse a cabo las medidas necesarias para impedir su identificación por voz o rostro.

La legislación española en cambio establece que, durante las declaraciones de los agentes en el proceso, estos podrán mantener su identidad encubierta siempre que así se acuerde mediante resolución judicial motivada. No obstante, nada se establece respecto a la excepcionalidad de esta medida o a la absoluta necesidad de llevar a cabo esa declaración.

Otro punto más en el que una legislación y otra encuentran diferencias entre sí está dado por el plazo de duración de la medida. Mientras que la española establece un máximo de seis meses prorrogables por periodos de igual duración, nada dice la argentina al respecto.

Por último, en la regulación española, a través de la LO 13/2015 del 5 de octubre de ese año, se incorpora la figura del agente encubierto informático que no está prevista en el caso de Argentina. [5]

En el artículo 6 de esa norma se prevé la posibilidad de que los funcionarios policiales actúen bajo identidad ficticia ya no únicamente de modo presencial, sino también en comunicaciones digitales (v. gr. chats, foros, canales cerrados, entre otros) con el fin de esclarecer delitos, autorizando para ello el intercambio o envio de archivos ilícitos.[6]

A modo de síntesis, ambas legislaciones –con algunas diferencias entre sí– establecen el modo de llevar a cabo la actuación del agente encubierto y la necesidad de una autorización judicial –o fiscal, con inmediato aviso al juez en el caso de España– que habilite la medida de modo fundado y con un análisis previo de razonabilidad y proporcionalidad. Regulan también la posibilidad de que los agentes encubiertos declaren en la etapa del proceso, sin perjuicio de los recaudos necesarios al llevar a cabo esa medida y eximen de responsabilidad a los funcionarios policiales que en el marco de su actuación como agentes lleven a cabo acciones ilícitas, siempre dentro de ciertos límites y criterios establecidos también legalmente.

5. Asencio Mellado, *José María, Justicia Penal y Nuevas formas de delincuencia*, México, Tirant Lo Blanch, 2015, p. 102.

6. Barona Vilar, Silvia, *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice*, Tirant lo Blanch, 2021.

Posturas jurisprudenciales y doctrinarias en torno a la figura del agente encubierto

Inconstitucionalidad del agente encubierto como técnica de investigación: posición minoritaria

Una parte importante de la doctrina considera que este instituto resulta violatorio de las normas procesales y garantías constitucionales que hacen al debido proceso legal (*v. gr.* 18 y 19 CN, art. 75 inc. 22 de la CN, arts. 8 y 21, de la CADH y 14 del PIDCP), principalmente en cuanto a que, mediante su utilización, pudiere afectarse la inviolabilidad del domicilio de la persona imputada, su derecho a la intimidad, su derecho de defensa en juicio, la garantía contra la autoincriminación forzada y el principio de inocencia.

Esta postura considera que la no revelación de identidad del agente por potenciales riesgos para su integridad, o el carácter extraordinario de su declaración durante el debate oral, implicaría una violación del derecho de defensa en juicio, la igualdad de armas en el proceso y el principio de contradicción. Concretamente, se cuestiona esta figura por impedir a la persona acusada en el marco de un proceso de controlar, interrogar y confrontar la prueba de cargo incorporada en su contra. O incluso, ante el desconocimiento de la identidad del agente, de verificar si hay cuestiones personales que pongan en juego su imparcialidad o fiabilidad.

Por otra parte, y respecto a la inviolabilidad del domicilio y al derecho de intimidad de la persona imputada, la postura que se inclina por la inconstitucionalidad de esta técnica considera que el agente encubierto, en su infiltración y viciando el consentimiento de quien tiene el derecho de permitir o excluir el ingreso de un tercero a su domicilio, lograría ingresar a su esfera de intimidad y acceder a conversaciones y documentos privados que no podrían recabarse de otro modo. Que no resultarían validos en términos legales por ser obtenidos mediante un engaño por el cual se vicia el consentimiento de la persona sospechosa que, de conocer la realidad de la situación, difícilmente hubiera permitido ese acceso resultando de ese modo un ingreso nulo.

En lo que a la garantía de la no autoincriminación respecta, el agente encubierto obtiene información producto de los dichos que la persona sospechosa manifestó frente a aquel, desconociendo su rol investigativo. Quienes critican esta figura entienden que, introducida esta información al juicio, colisiona con el principio de inocencia y la garantía de no autoincriminación forzada que protege a persona imputada a no ser inducida a declarar contra sí misma de manera coactiva o engañosa, sino únicamente, de forma libre y voluntaria.

Constitucionalidad de la técnica investigativa: postura mayoritaria

Si bien la utilización de la figura del agente encubierto no resulta pacifica a nivel doctrinario ni jurisprudencial, nuestra Corte Suprema de Justicia ha optado por receptar y aplicar esta técnica de investigación frente a la necesidad de llevar a cabo procesos penales efectivos contra delitos que debido a su complejidad resultan difíciles de ser descubiertos y sancionados con las medidas de prueba tradicionales.

Incluso previo a la sanción Ley Nº 27319, nuestro máximo Tribunal estableció la validez del agente encubierto. En el caso "Fiscal c. Fernández, Víctor Hugo"[7] valoró que aquella no es en sí misma violatoria de garantías constitucionales siempre y cuando se mantenga dentro de los principios rectores del Estado de derecho y no hubiese creado o instigado de manera directa la ofensa criminal en cabeza del delincuente, actuando como un agente provocador.

El fundamento de la Corte en este caso fue que no se ven corrompidas las garantías constitucionales en juego en tanto el ocultamiento de la verdadera identidad policial sólo tiene por objeto tomar conocimiento de un hecho que es realizado libremente por la persona imputada.

Sumado a la necesidad de los Estados de implementar técnicas especiales de investigación a los fines de desentrañar ciertos delitos complejos que sólo son susceptibles de ser descubiertos y probados si los órganos encargados de la prevención logran ser admitidos en el círculo de intimidad donde se desarrollan.

7. CSJN, Fallos: 313-2:1305, "Fiscal c/ Fernández, Víctor Hugo s/ av. infracción Ley N° 2077", 11/12/1990.

Un paso más hacia allá en la era digital: la utilización de IA en la figura del agente encubierto. El caso *"Sweetie"*

A finales del año 2013 en Holanda, la ONG *Terre des Hommes* desarrolló un proyecto denominado "*Sweetie*" con el fin de identificar personas que buscan menores en foros, chats y otros lugares del ciberespacio. *Sweetie* es un robot con la apariencia de una niña de origen filipino de 10 años creada a través de técnicas de animación avanzadas que captan los movimientos y la voz de una persona real (de una niña) que fue utilizada por un agente de dicha ONG para recoger la información de personas que la iban contactando y poniéndola en conocimiento de las autoridades de distintos países.

En un primer momento se trató de un avatar detrás del cual operaban agentes que eran quienes trataban con los pedófilos que se contactaban. Sin embargo, un problema que planteó este mecanismo fue de escalabilidad: ante el gran caudal de personas que la contactaban, los agentes que operaban tras ella no lograban mantener conversaciones simultaneas con todos. En ese escenario se creó la segunda versión: "*Sweetie* 2.0" en la que con ayuda de inteligencia artificial se logró resolver esa cuestión.

Actualmente, para este *bot* es posible interactuar con posibles pedófilos e incluso mandar fotos o mostrarse por *webcam*; realizando la actividad de un funcionario encubierto informático, pero superando en demasía el número de conversaciones en tiempo real que podría realizar un agente humano.

Análisis acerca de la viabilidad de utilizar un agente encubierto con inteligencia artificial en el estado legislativo actual

En el estado legislativo actual, la figura del agente encubierto dotado de inteligencia artificial no se encuentra expresamente regulado. De allí que su eventual utilización exige replantearse ciertas cuestiones, no

sólo a la luz de las garantías constitucionales, sino a si resulta o no posible su aplicación técnica con las previsiones legales actuales.

Principio de legalidad

Toda medida de investigación que implique una injerencia estatal en un derecho constitucional está gobernada por la regla de la taxatividad legal y será necesario que una norma que la autorice y establezca los requisitos a cumplir.

Dicho de otro modo, toda acción por parte del estado que conlleve la restricción de un derecho protegido constitucionalmente requiere a los fines de su validez una decisión legislativa que así lo determine.

Actualmente, nuestra legislación faculta expresamente a los órganos encargados de la investigación a utilizar agentes encubiertos en determinados casos y bajo ciertos requisitos, entre los que incluye que la actuación sea llevada a cabo por un "funcionario policial". Esto a simple vista parecería descartar la posibilidad de que la infiltración sea llevada a cabo por un robot que no reúne estrictamente la calidad de agente policial.

A los fines de sortear este obstáculo hay quienes podrían sostener que la técnica de investigación es aplicada por las fuerzas de seguridad. De modo que, si bien en un sentido estricto no es un funcionario policial quien se infiltra, sí lo hace de manera indirecta a través del robot.

Pero como postura contraria se podría argumentar que no necesariamente es correcto considerar que es un funcionario policial quien indirectamente ejecuta la medida. No hay dudas que cuando un agente de manera personal se introduce en una red delictiva, es él quien se está infiltrando. En estos casos, en cambio, también sería posible considerar que el robot lo ejecuta el juez a través de su orden, el fiscal a través de su pedido, o incluso el Ministerio de Justicia a través de su control.

Otro obstáculo que esta asimilación encuentra es que aun de considerar que es efectivamente un funcionario policial quien ejecuta la infiltración, esto conllevaría a una aplicación análoga de la normativa vigente, vedado expresamente en materia penal por el principio de legalidad.[8]

8. CSJN, Fallos: 314:1451, "Martinez Perea, Jerónimo s/contrabando", 12/11/1991.

Si bien es cierto que en nuestro sistema procesal rige el principio de libertad probatoria, lo que implica que los hechos y circunstancias pueden ser acreditados con cualquier medio en la medida que sean idóneos para esclarecer el hecho objeto de investigación, no debe perderse de vista en este punto que un límite formal a este principio es la CN y que toda medida de prueba que implique una injerencia estatal solo puede llevarse a cabo respetando las normas constitucionales. Lo que dicho de otro modo trae aparejado que la restricción a los derechos no pueda ser realizado en cualquier caso o mediante la aplicación análoga de otras medidas, sino únicamente a través de una ley u orden de autoridad competente en virtud del principio de legalidad penal.

Este parecería ser el primer obstáculo que encuentra la utilización de un agente encubierto con inteligencia artificial en nuestro estado legislativo actual, por la ausencia de una norma que de manera exprese regule este supuesto y habilite una injerencia del Estado en la esfera de intimidad y derechos de la persona sospechosa de cometer un delito.

Derecho de defensa y posibilidad de interrogar testigos

Nuestro modelo de enjuiciamiento actual exige de la existencia de un debate oral y público a través del contradictorio entre las partes, como presupuesto para que el tribunal de juicio adopte la decisión final. Ese confronte es la vía exigida por el legislador para que la parte acusadora tenga la posibilidad de demostrar la imputación y la persona acusada junto con su abogado pueda ejercer su defensa.

Dentro de este derecho a defenderse se incluye la posibilidad de la persona a controlar la prueba de cargo, que no se reduce a la simple observación de los elementos de juicio presentados en su contra sino a la posibilidad efectiva de actuar sobre ellos.

El artículo 8 de la CADH establece, entre las garantías mínimas del proceso, el derecho de la defensa de interrogar a los testigos presentados en el tribunal y de obtener la comparecencia de todas aquellas personas que puedan arrojar luz sobre los hechos. La CIDH ha sostenido, con cita de jurisprudencia del TEDH[9] que dentro de las prerrogativas

9. Corte Europea de Derechos Humanos en el "Caso de Barberá, Messegué y Jabardo", 6 de diciembre de 1998, Ser. A N° 146, párr. 78 y Corte Europea de Derechos Humanos,

que deben concederse a la persona acusada está la de examinar los testigos en su contra y a su favor con el objeto de ejercer su defensa.[10]

Análogamente, debe otorgarse al acusado acceso a los documentos y demás pruebas en posesión y control de las autoridades, necesarias para la preparación de su caso.[11]

La CSJN ha recordado en esta línea que es violatorio del derecho de defensa la utilización de una base probatoria sobre la cual no se haya tenido siquiera la posibilidad de controlar dicha prueba.[12] En igual sentido, se estableció con cita al TEDH[13] que el derecho de examinar exige que el imputado haya tenido al menos una oportunidad adecuada y apropiada para desafiar y cuestionar a un testigo o cualquiera que hubiera hecho declaraciones en su contra.

En definitiva, toda prueba testimonial debe ser sometida al contradictorio para gozar de plena validez y eficacia, posibilitando el control de la defensa y respetando de ese modo el derecho de defensa de la persona imputada.

En resguardo de esta garantía, la norma prevé expresamente la posibilidad de que el funcionario policial que actúa como agente encubierto declare en la etapa de juicio. Esto sin perjuicio de aclarar la excepcionalidad de la medida, y los reparos a llevarse a cabo para proteger su integridad física.[14] En estos casos, existe una posibilidad fáctica de que, en miras a evitar una vulneración al derecho de defensa, la persona que lleva adelante la infiltración pueda declarar en juicio y ser interrogada a fin de incorporar su actuación como prueba válida de cargo.

Caso Bönisch, 6 de mayo de 1985, Serie A, N° 92, párr. 32, citado en el "Caso Castillo Petruzzi y otros", nota 55 supra, párrs. 153-154.

10. Corte IDH, "Castillo Petruzzi y otros vs. Perú", 30/05/1999.

11. Comité IDH, Observación General N° 13, Principios Básicos de la ONU sobre la Función de los Abogados, Nota N° 545 supra, párr. 9; Nota N° 589 supra, artículo 21: "Las autoridades competentes tienen la obligación de velar por que los abogados tengan acceso a la información, los archivos y documentos pertinentes que estén en su poder o bajo su control con antelación suficiente para que puedan prestar a sus clientes una asistencia jurídica eficaz".

12. CSJN, Fallos: 329:5556, "Benítez, Aníbal Leonel s/ lesiones graves", 11/12/1990.

13. TEDH, "Säidi vs. Francia", Serie A, N° 261-C, 20/09/1993.

14. La doctrina no es pacífica en este punto. Una primera postura entiende que el agente opera como un mero instrumento de investigación y por ende no es necesario que declare en juicio. La posición contraria sostiene que al ser una herramienta probatoria, su comparecencia al debate es fundamental.

No pareciera ocurrir lo mismo en el caso de que quien actúa como agente encubierto no sea una persona humana, sino un robot dotado de inteligencia artificial. La propia naturaleza del agente –no humano– y la imposibilidad de ser sometido a un interrogatorio podría entrar en pugna con el derecho de defensa y al propio principio contradictorio característico de esa etapa.

Una posible solución en estos casos podría ser considerar la actuación llevada adelante por parte del robot, no como una prueba de carácter testimonial donde se alza la necesidad de un interrogatorio, sino como una de tipo documental al que la defensa tenga acceso. Para eso, una ley debería establecer el procedimiento a seguir para que toda su actuación llevada a cabo de manera digital sea debidamente registrada, cuidando la cadena de custodia, y permitiendo su incorporación al proceso como un documento controlable por las partes.

La ley española en su sección V regula las cuestiones relativas a la interceptación de comunicaciones telefónicas y telemáticas. En su capítulo 1, artículo 588 ter prevé sus presupuestos entre los que incluye a los delitos cometidos a través de instrumentos informáticos o de cualquier otra tecnología de la información o la comunicación o servicio de comunicación y establece que una vez expirada la medida de intervención, las partes recibirán copia de las grabaciones y de las trascripciones realizadas.

Un paralelismo entre este supuesto y el nuestro de análisis permite advertir que parecería viable una regulación en la que al autorizar la intervención del agente encubierto con inteligencia artificial se autorice además la registración de toda su actuación y todos sus contactos con la persona de sospechosa. De modo tal que, una vez finalizada la intervención, el registro documentado que quede de ello sea accesible y controlable para las partes. Las cuales ya no confrontarán la misma como una de carácter testimonial, sino que lo harán a través de la verificación de que estén dados los requisitos legales y de que la actuación del agente haya sido llevada a cabo dentro de los límites legales.

Eximición de responsabilidad por los delitos cometidos en el marco de la actuación del agente

La legislación argentina al igual que la española, aunque con ciertas diferencias entre sí, prevé para la figura del agente encubierto una eximición de responsabilidad. Si en el marco de su actuación comete acciones ilícitas necesarias para lograr un fin que le fue asignado, la consecuencia es que ellas no deriven en una sanción penal en su contra.

Así, el artículo 282 bis de la ley de enjuiciamiento criminal español establece

> … el agente encubierto estará exento de responsabilidad criminal por aquellas actuaciones que sean consecuencia necesaria del desarrollo de la investigación, siempre que guarden la debida proporcionalidad con la finalidad de la misma y no constituyan una provocación al delito.

En tanto la legislación argentina, en el artículo 9 de la Ley Nº 27319, establece

> … no será punible el agente encubierto o el agente revelador que como consecuencia necesaria del desarrollo de la actuación encomendada, se hubiese visto compelido a incurrir en un delito, siempre que este no implique poner en peligro cierto la vida o la integridad psíquica o física de una persona o la imposición de un grave sufrimiento físico o moral a otro.

De una y otra regulación se desprende entonces la existencia de una eximente de responsabilidad para aquellos delitos cometidos i) en el marco de su actuación como agente, ii) que resulten necesarias para el desarrollo de su actuación, iii) que no conlleven, en el caso de la legislación española, una desproporción con la finalidad o resulten una provocación al delito. O que no impliquen, en el caso de la legislación argentina, una lesión a la vida o integridad física de un tercero.

Esta ausencia de punibilidad ampara exclusivamente al funcionario policial que lleva a cabo la infiltración y durante el marco de su actuación. No resulta extensible ni a otros sujetos, ni al mismo agente fuera de ese ámbito.

No ofrece dificultades de interpretación el caso del agente encubierto tradicional, como tampoco la figura del agente encubierto-informático, pues en un caso y en otro quien cumple ese rol, sea de

manera presencial o virtual, es una persona humana. Y será ella quien resulte eximida de responsabilidad penal por los ilícitos llevados a cabo en el marco de su actuación.

Este panorama no parece tan claro cuando se analiza esta figura con la utilización de inteligencia artificial. En términos estrictamente literales, el "agente" pasible de eximición de responsabilidad, sería el robot que, programado a tales fines, lleva a cabo las acciones necesarias –entre ellas, puede ocurrir que algunas delictivas– para infiltrarse en círculos cerrados de criminalidad y obtener información.

Si las conductas del robot se encuadran en todo momento en un marco legal parecería que el caso no ofrece mayores dificultades. Sin embargo, podría devenir en un análisis más complejo sí con su actuación el robot cometería alguna acción ilícita.

De tratarse de aquellas amparadas por la eximición de punibilidad, el debate simplemente giraría en torno a cuestionarse a quien se estaría eximiendo en el caso: al programador del robot, a la fuerza policial que lo utiliza, a las fuerzas estatales que lo ordenan, que lo controlan, etcétera. No obstante, se trate de una u otra figura, la consecuencia práctica devendría de todas formas en una eximición de responsabilidad por estar así previsto legalmente.

El mayor problema parecería aparecer en un caso donde el robot se excediera en sus acciones y llevara a cabo una conducta delictiva fuera de los permisos legales, por ejemplo –tomando como base la legislación argentina–, cometiendo un hecho que dañe la integridad física o psíquica de un tercero.

No estando amparada esta acción por la cláusula de eximición de responsabilidad, cabría preguntarse quién resulta penalmente responsable por los daños causados por el robot.

En un caso de intervención de un agente encubierto tradicional, así como es él el acreedor de la eximición de punibilidad, también lo es de la responsabilidad penal si su actuación se extralimita de los permisos legales. Pero esto que es claro en la figura tradicional, no pareciera tan sencillo en el caso de un agente encubierto con inteligencia artificial.

Esto desde ya forma parte de una vigente discusión que se suscita alrededor de cuál es la responsabilidad que cabe a los fabricantes de

las máquinas, y que excede por mucho al objeto de este trabajo.[15] Lo que en definitiva puede sostenerse es que la regulación actual no parecería poder brindar respuestas legales en este punto acerca de quién es acreedor de la eximición de responsabilidad que prevé la norma, o quien es la persona humana responsable por las acciones del robot en caso de que éste se exceda de los límites legales establecidos.

Delito imposible o tentativa inidónea

Las acciones de tentativa tienen diferentes grados de idoneidad para consumar el delito y resulta necesario determinar cuáles son los límites típicos que separan a aquellas conductas que afectan el bien jurídico y por ende habilitan la punición estatal de aquellas otras que por ser inidóneas determinan una imposibilidad absoluta de consumar el delito y no encuadran dentro de esta categoría.

Explica al respecto Zaffaroni que existen diferentes supuestos: el de tentativa aparente, cuando no hay tentativa por ausencia de tipicidad que se presenta por la falta de algún elemento del tipo legal o sistemático (ausencia de tipo en sentido estricto), o bien cuando falta la cosa elegida como medio (atipicidad por falta de medio) o cuando *ex ante* el medio elegido carece de cualquier idoneidad para consumar el delito (atipicidad por falta de medio idóneo). El segundo supuesto que plantea el autor y que es el que aquí interesa es el de delito imposible. Entran dentro de esta categoría aquellos casos en que *ex ante* el medio fue idóneo y hubo peligro, no obstante, *ex post*, por la forma inadecuada en que se usó el medio, por un grave defecto de este, un accidente en el objeto o por una previa neutralización del peligro, se determina una imposibilidad absoluta de consumar el ilícito.[16]

La doctrina argentina identifica este último supuesto de delito imposible previsto en el artículo 44 de Código Penal con la tentativa inidónea. El fundamento de esta punición se basa en que, si bien existió un peligro por la idoneidad del medio, se da en el caso un injusto

15. Danesi, Cecilia, "¿Quién responde por los daños ocasionados por los robots?", en *Revista de responsabilidad civil y seguros: publicación mensual de doctrina, jurisprudencia y legislación*, 2018.

16. Zaffaroni, Eugenio Raúl, *Derecho Penal Parte General*, Buenos Aires, Ediar, 2014, pp. 832-833.

de menor entidad por comprobarse ex post que el delito era absolutamente imposible de ser consumado.

Con un criterio similar, Mir Puig desarrolla que hay delito imposible o tentativa inidónea (entendiendo ambos como sinónimos) cuando por la inidoneidad del objeto, medios o sujeto, no era posible llegar a la consumación del delito. Diferencia aquellas acciones que en un principio eran capaces de consumar el ilícito, aunque *ex post* no lo logren, de aquellas otras que aparecen como incapaces de lograrlo desde un primer momento. Solo a éstas últimas las incluye dentro de la categoría de delito imposible.

En cuanto al fundamento de su punición, explica el autor español que la tentativa inidónea es peligrosa *ex ante* teniendo en cuenta que, para un espectador objetivo situado en el lugar del hecho, la situación hubiera parecido capaz de consumar el delito. Por eso, esa apariencia de idoneidad ex ante crea un peligro abstracto, a diferencia del peligro concreto de una tentativa idónea. Diferente es el caso en que ya desde un inicio, para el propio espectador, se carezca de toda base de posibilidad para la consumación y no estaríamos ya frente a una tentativa inidónea sino una irreal.[17]

Aplicado esto a la utilización de un agente encubierto se podría plantear un escenario en el que las conductas llevadas a cabo por la persona sospechosa resulten en definitiva delitos imposibles.

No sería correcto incluir dentro de esta posibilidad a todas las acciones llevadas a cabo por el autor, pues hay algunas, como la distribución toda representación de un menor dedicado a actividades sexuales explícitas con fines sexuales que sin importar que del otro lado de la pantalla se encuentre un humano o un robot, podría llegar a considerarse consumado el delito.

Distinto parecería ser el caso en el que el tipo legal exigiese ciertas características para la consumación que de no darse se vería frustrada su posibilidad de punición.

Podría por ejemplo mencionarse aquellos casos en que la utilización de esta herramienta se realiza con el fin de identificar pedófilos que interactúan en la red con menores de edad. Si el delito que

17. Mir Puig, Santiago, *Derecho Penal, Parte General*, Buenos Aires-Montevideo, Editorial BdeF, 2018, pp. 364-366.

se intenta perseguir y prevenir a través de este mecanismo es el de *grooming*, la intervención del agente encubierto podría implicar la eliminación de uno de los elementos del tipo penal convirtiendo en atípica la conducta del autor.

Si la figura de *grooming* exige para su configuración que el contacto sea mantenido con un menor de edad, la utilización de un agente encubierto con inteligencia artificial (o incluso sin ella, pero con un adulto del otro lado de la pantalla), podría conllevar a la atipicidad del delito por no cumplirse las características en el sujeto pasivo que este exige.

Este pareciera ser un nuevo obstáculo que encuentra la utilización de un agente encubierto, ya no solo únicamente en la utilización inteligencia artificial sino también en su versión virtual. Pues en uno y otro supuesto quien en definitiva estaría interactuando con el autor sería un robot o un funcionario policial en forma digital, pero no un menor como exige el tipo penal.

De este modo se advierte que la figura del agente encubierto no parecería la adecuada para dar inicio a un proceso en contra de posibles autores del delito de *grooming* pues, por los propios requisitos normativos, ya desde el inicio la conducta devendría en atípica. Sin embargo, sí podría resultar un elemento eficaz para recolectar material probatorio respecto de aquellos potenciales autores que se encuentran siendo investigados por contactos mantenidos con menores de carne y hueso.

Palabras finales

Los desafíos que los avances tecnológicos plantean para el procedimiento penal y las garantías constitucionales son innegables. Los beneficios que estos medios pueden traer para la vida social en general y para el derecho en particular, no queda exentos de ser utilizados para actividades delictuales de gran complejidad que impone a los estados la necesidad de contar con los medios y respuestas eficaces para combatirlas.

En este contexto surgen las técnicas especiales de investigación y entre ellas el agente encubierto que, no sin ciertos reparos y posturas en contra, viene desde hace ya tiempo siendo utilizado para intentar combatir los delitos complejos a los que los Estados se enfrentan.

En un escenario ya poco pacífico, irrumpe la inteligencia artificial para dar comienzo a una nueva forma de pensar y utilizar las técnicas de investigación tradicionales.

La figura del agente encubierto con inteligencia artificial, si bien aún en etapa embrionaria a nivel mundial, se encuentra en alza en distintos Estados por los resultados exitosos que su intervención presentó.

Esto necesariamente nos obliga a plantearnos ciertas cuestiones relativas a esa herramienta y a la viabilidad de su aplicación a la luz de las garantías constitucionales.

En la regulación tradicional se prevén cuestiones como la intervención de un funcionario policial, la eximición de responsabilidad o la posibilidad de que el agente declare en la etapa de juicio, que están pensadas para ser aplicadas frente a un funcionario humano, pero no un robot.

Esto no se reduce a una mera cuestión técnica, sino que detrás de estas cuestiones aparecen garantías vinculadas como el propio principio de legalidad o el derecho a interrogar a un testigo, que necesariamente exigen una revisión y adecuación de la normativa que hoy parecería insuficiente.

Como fue analizado a lo largo de este artículo, sin perjuicio de las posturas contrarías que la figura del agente encubierto presenta y que no fueron objeto de este trabajo, lo cierto es que su aplicación con inteligencia artificial no pareciera posible sin una modificación en la norma que no solo autorice de manera expresa su utilización sino que además dé respuesta a las cuestiones que fueron aquí analizadas y que no logran ser resueltas con la regulación actual.

Es esa la única forma de lograr un equilibrio entre el legítimo interés de los Estados en la investigación de los delitos, y el respeto de las garantías constitucionales que rigen en un Estado de derecho.

Inteligencia Artificial en las sentencias de la justicia criminal

María Catalina Rangugni*

La máquina había captado la forma de la narración de Poe y le había cambiado la anécdota, por lo tanto, era cuestión de programarla con un conjunto variable de núcleos narrativos y dejarla trabajar. La clave, dijo Macedonio, es que aprende a medida que narra. Aprender quiere decir que recuerda lo que ya ha hecho y tiene cada vez más experiencia. No hará necesariamente historias cada vez más lindas, pero sabrá las historias que ha hecho y quizá termine por construirles una trama común.
Ricardo Piglia, *La ciudad ausente*

Introducción

La era contemporánea, marcada por avances tecnológicos de proporciones sin precedentes, se encuentra inmersa en una revolución[1] que está redefiniendo la forma en que interactuamos entre nosotros, con nuestra comunidad, con el mundo y, de manera particular, con la justicia. Una de las aristas que genera grandes niveles tanto de curiosidad académica como de controversia en este contexto es el debate sobre los "jueces robots" o sistemas de inteligencia artificial (IA) diseñados para tomar decisiones judiciales.

Ese tipo de implementación de herramientas de IA obliga a hacerse algunas preguntas esenciales sobre la equidad, la imparcialidad y la naturaleza humana en el sistema legal. A medida que la tecnología avanza y se integra cada vez más en nuestras instituciones, es imperativo examinar en profundidad los beneficios y los desafíos que

* Abogada (UBA). Magíster en Derecho Penal y Procesal Penal (Univ. San Andres) y en Justicia Criminal (Queen Mary Univ. London).

1. Para indagar en esta idea, basada en lo que Sherry Turkle ha llamado "el horizonte robótico", puede consultarse el libro Turkle, Sherry, *Alone Together: Why We Expect More from Technology and Less from Each Other*, Nueva York, Basic Books, 2012.

acompañan esta creciente influencia de la inteligencia artificial en uno de los pilares esenciales de la sociedad: el sistema de justicia. En este ensayo, exploraré sucintamente las características, el funcionamiento y las implicaciones éticas de los jueces robots, evaluando tanto sus ventajas potenciales como las preocupaciones críticas que se plantean para la búsqueda continua de un sistema legal equitativo y humano.

¿Cómo puede intervenir un modelo algorítmico en una sentencia?

En la actualidad es usual escuchar hablar del concepto de "jueces robots" que se ocupan de juzgar y redactar sentencias. Ahora bien, en concreto, ¿de qué manera se puede materializar la intervención de un modelo algorítmico en un juicio o en la redacción de una sentencia?

Si bien podrían diagramarse infinitos escenarios con distintas escalas de participación de *softwares* de IA en decisiones judiciales, a los efectos de este artículo resulta útil distinguir dos categorías: El reemplazo total de los jueces por un algoritmo y la utilización de la IA para asistir al juez en sus decisiones y en la redacción de las sentencias.

El primer escenario, en el que la figura del juez humano desaparece por completo, supone que el rol de los operadores judiciales se limite al de "*data entry*", y que sea el sistema computacional el que pondere las características propias de la materialidad, los principios del derecho aplicables y la solución "justa" para el caso concreto. Lo cierto es que no es común leer autores que propicien este tipo de aplicación de la tecnología en la justicia criminal, dado que, más allá del interés que puede suscitar una hipótesis semejante desde el ámbito académico, en mi opinión, no pareciera ser un planteo realista –al menos en lo inmediato– frente al actual estado de la ciencia.

Sin embargo, una alternativa que podría otorgar algún grado de viabilidad a ese primer planteo sería emular lo que ocurre en algunas jurisdicciones argentinas con el juicio por jurado, o con los tribunales colegiados, es decir, dar la opción a las partes para que puedan elegir si prefieren una sentencia confeccionada por la inteligencia artificial por encima del juicio efectuado por un juez o un tribunal de personas humanas. Algunos autores proponen que soluciones de este tipo sean

acompañadas de la conformidad de ambas partes aceptando lo decidido por el sistema de IA para que la sentencia adquiera firmeza.[2]

En segundo lugar, el rol que puede darse a modelos de IA en la redacción de las sentencias es el de asistir la decisión de los jueces. Este abordaje es notoriamente más usual en cuanto a la aplicación concreta de IA en la justicia y a la hora de pensar el diseño de los prototipos que aún se encuentran en desarrollo. La idea central es que el juez consulte la recomendación del *software* diseñado a ese efecto y pueda tenerlo en cuenta al momento de resolver pero sin que ello sea vinculante para su decisión. Este es el enfoque que utilizan los sistemas que se han desarrollado en los últimos años, tales como COMPAS, en Estados Unidos, Sócrates, en Brasil y el resto de ejemplos que enumeraré más adelante.[3]

Utilizadas de esta manera, las herramientas digitales pueden contribuir en distintos aspectos de las decisiones judiciales, ya sea sugiriendo soluciones a algunos aspectos que requieren cálculos de ponderación (como pueden ser pronósticos de reinserción, mensuración de la pena, etc.), colaborando en la revisión de planteos recursivos o proponiendo principios rectores para la unificación de criterios.

Un enfoque de estas características también puede ser planteado de manera que sean los jueces quienes se ocupen íntegramente de elaborar sus fallos, sin la asistencia de la IA antes o durante el proceso de toma de decisión, sino después de él, para revisar defectos o errores, tales como inequidades injustificadas, discriminación, vulneración de principios constitucionales o incongruencias. Esto se conoce como Aprendizaje Algorítmico del Humano (HAL por sus siglas en inglés),[4] debido a que el algoritmo aprende de las decisiones humanas.

En ese tren, instancias de revisión (Tribunales Superiores) pueden valerse del asesoramiento de modelos algorítmicos para analizar arbitrariedades graves en los fallos que les toca revisar e incluso para elaborar lineamientos unificados para los tribunales de primera instancia.

Dicho esto, nada obsta a la construcción de sistemas híbridos que, por ejemplo, incorporen fallos íntegramente confeccionados por

2. Schearze, Mathis; Roberts, Julian, "Reconciling Artificial and Human Intelligence: Supplementing Not Supplanting the Sentencing Judge". Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3872766 [Fecha de consulta: 11/03/2024].

3. Ver en este artículo "Modelos Aplicados".

4. Schearze, Mathis; Roberts, Julian, *op. cit.*

IA y otros mediante labor humana, pudiendo determinar en qué casos corresponde cada solución según la gravedad, la complejidad de la prueba, el volumen del expediente, la fecha de prescripción, etcétera. Otra figura híbrida podría ser la de un juez "data entry" que revise las fuentes, la información relevante y los valores asignados por el modelo algorítmico para la decisión final del caso o, por otro lado, la de un juez que revise y haga las consideraciones correspondientes al caso y luego ingrese la información en el sistema para que tome la decisión final.

Lo cierto es que la idea de "Jueces Robots", tal como se presenta en un modelo de total reemplazo de los jueces humanos es menos probable de lo que la vorágine propia de la novedad pareciera indicar, especialmente porque el razonamiento legal no es un procedimiento puramente lógico.

Las sociedades modernas se han volcado y se siguen desarrollando progresivamente hacia el racionalismo absoluto, buscando que los problemas de los hombres y las mujeres modernas se resuelvan, cada vez más, a través de fórmulas matemáticas que garanticen altas probabilidades de éxito. Parece que hubiéramos trasladado nuestra fe a oráculos virtuales a los que les pedimos respuestas sobre nuestra propia existencia. Sin embargo, el derecho –especialmente el criminal– es una ciencia humanística, social, atada a la dinámica de las comunidades en las que opera.

No debe confundirse el carácter normativo de las leyes con la linealidad de estructuras lógicas rígidas como las que necesariamente gobiernan cualquier razonamiento matemático. Por eso, si bien el cálculo de riesgos y el análisis de precedentes por parte de la IA puede agregar elementos de valor al juez para decidir sobre un caso, la tarea interpretativa requiere esencialmente de características humanas como el sentido común, moral, social y cultural.[5]

La transparencia en las sentencias judiciales

Ahora bien, en lo que a la ética propia de los sistemas de IA respecta, importa a los fines de este ensayo poner el foco en la transparencia y en la ecuanimidad, para lo cual el primer obstáculo a sortear es la

5. Susskind, Richard, *The future of law*, Oxford, Oxford University Press, 1986, p. 133.

tensión que puede presentarse entre la precisión en las predicciones que puede llegar a alcanzar un algoritmo mediante su complejización y la pérdida de transparencia en ese proceso.

En términos de transparencia, los sistemas de inteligencia artificial utilizados para asistir la redacción de sentencias tienen un lado luminoso y un reverso oscuro. Es decir, existen características de la IA que pueden potenciar la translucidez de una decisión judicial, pero también existe el riesgo de que ocurra lo contrario.

Para entender esta dualidad, resulta útil repasar las razones por las que un algoritmo puede nublarse de "opacidad":

a. Legal: Generalmente los modelos algorítmicos son desarrollados por entes privados que, para protegerse comercialmente de la competencia, gozan del secreto societario respecto de sus productos. En tal sentido, una compañía para mantenerse en la delantera de sus competidores puede pretender no revelar el detalle de las fuentes y códigos que utiliza, cómo es que fue diseñado, programado y/o desarrollado el *software* que ofrece[6] (por ejemplo, caso "Loomis v. Wisconsin").
b. Técnica: Cada vez más se habla del problema de "cajas negras" que producen los modelos de redes neuronales artificiales que funcionan con capas.[7] En prieta síntesis, algunos sistemas de IA, por el modo en que operan, impiden que pueda conocerse, en todo o en parte, cómo fue el procedimiento paso por paso por el que arribaron al resultado brindado. Esto hace que ciertas fuentes, *inputs* o valores sean ignorados por quien utiliza el sistema –en algunos casos hasta para quien diseña, programa o desarrolla el sistema, dado que más allá de las directivas que pueden darle al modelo algorítmico sus creadores, la idea de que ciertos modelos pueden "aprender por sí mismos" implica que en determinadas ocasiones la capacidad

6. Wisser, Leah, "Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing", en *American Criminal Law Review*, Vol. 56:1811, Georgetown, Georgetown University Law Center, 2019.

7. Este problema fue abordado con mayor extensión en mi artículo anterior: Rangugni, María Catalina, "Algoritmos predictivos para optimizar las cárceles: ¿puede calcularse el riesgo de reincidencia?", en Eidem, Matías E.; Kleiman, Hernán (dirs.), *Nuevas visiones del derecho penal*, Buenos Aires, Ad-Hoc, 2023.

de síntesis de la máquina impida conocer su razonamiento cabalmente.[8]

c. Analfabetismo algorítmico: También puede darse el supuesto en que legal y técnicamente el algoritmo sea transparente pero que el mecanismo por el cual arribó a un resultado determinado no sea fácilmente desglosable para ciudadanos que no se encuentran empapados de la temática y que, en definitiva, son en la mayoría de casos quienes terminan siendo objeto de las sentencias.

Dicho esto, y con la intención de poner luz sobre las circunstancias a atender para favorecer la transparencia de los sistemas algorítmico, cabe preguntarse: ¿Cuál es el objeto de opacidad del algoritmo?

d. El código fuente:[9] Se trata de las indicaciones, declaraciones y funciones que los modelos de IA reciben para cumplir sus funciones, escritas en un lenguaje de programación determinado, que para que sea ejecutable debe ser traducido por algún programa que lo traduce a código binario (combinaciones de 1 y 0 que es el lenguaje que utiliza el *hardware* de una computadora). Es decir, son las fórmulas lógicas que utiliza el sistema para operar.

Si bien parece evidente que la transparencia del código fuente es esencial para emitir una sentencia respetuosa de la garantía del debido proceso penal, pues forma parte del derecho del imputado a controlar la prueba que se produce en su contra, lo cierto es que las empresas desarrolladoras de este tipo de *software* en distintos lugares del mundo reclaman que sobre ellos se respete el secreto empresarial, dado que exponer sus fórmulas podría derivar en un importante perjuicio para sus negocios. El argumento para sostener esta posición es que la innovación en herramientas para la Justicia está dada mayormente –o con la mayor efectividad– por parte del sector privado, por lo que quitarle incentivos a ese sector iría en desmedro del desarrollo de nuevas tecnologías de mejores resultados.[10]

8. Ídem.

9. Raymond, Darrell, "Reading source code", en *Proceedings of the 1991 conference of the Centre for Advanced Studies on Collaborative research*, Toronto, IBM Press, octubre de 1991.

10. Riquert, Marcelo, *Inteligencia artificial y derecho penal*, Buenos Aires, Ediar, 2022, pp. 101-131.

El valor otorgado a los *inputs* para obtener el *output:* Los sistemas de redes neuronales artificiales, en particular los de evaluación de riesgos (o RATs),[11] otorgan distintos valores a las variables que se introducen para hacer una predicción o tomar una decisión. Esta ponderación de factores puede ser muy relevante para el resultado que arroja el *software* y es particularmente compleja cuando se trata de aplicar IA a sentencias criminales.

Pongamos el caso del principio de proporcionalidad como ejemplo: Aplicar este principio en una sentencia supone ponderar factores que no son idénticos en todos los casos, por ejemplo la magnitud del daño infligido, el riesgo de reincidencia, qué relevancia se da al remordimiento del acusado o a la premeditación. No está claro de qué manera la IA podría contribuir a desentrañar el significado jurídico de los elementos que componen una sentencia o cómo puede programarse la IA para distinguir estos valores en aras del principio de proporcionalidad de cada caso. Sin embargo, el uso de modelos de *machine learning* podría ser útil para revisar *ex post* si los tribunales actúan de conformidad con ese principio.[12]

Cuál sea el objeto puntual de opacidad dependerá, como es inferible, de las causas antes mencionadas. Es decir, puede ser que por razones legales de secreto societario una empresa no revele las fuentes utilizadas por el algoritmo para operar, puede ser que por ese motivo decida revelar las fuentes mas no los valores otorgados para obtener el resultado, o puede ser que no revele ninguno de los dos datos. En esa misma línea, es posible que la estructura del algoritmo alcance un nivel de complejidad tal que alguno de esos factores o ambos no puedan ser conocidos debido a razones técnicas.

Para profundizar en este aspecto, y vincularlo directamente con la materia que nos ocupa, Jesper Ryberg analiza la relevancia de la opacidad de los algoritmos utilizados para producir sentencias, en función de su tensión con la posibilidad de fundamentar claramente las decisiones judiciales.[13]

11. Por su sigla en inglés: Risk Assessment Tools (RATs).

12. Chiao, Vincent, "Predicting Proportionality: The Case for Algorithmic Sentencing", en *Criminal Justice Ethics*, Vol. 37, Nueva York, Taylor and Francis, 2018.

13. Ryberg, Jesper, "Sentencing and algorithmic transparency", en Ryberg, Jesper; Roberts, Julian V. (Eds.), *Sentencing and Artificial Intelligence (Studies in Penal Theory and Philosophy)*, Nueva York, Oxford University Press, 2022.

La importancia de la fundamentación clara de las decisiones judiciales radica, en primer lugar, en la considerable mejora que esto produce en la calidad de las decisiones. Es de suponer que los jueces, al verse obligados a fundar en detalle sus decisiones, se esfuercen en que estén sostenidas por razones sólidas y esto puede ayudar a reducir sesgos o prejuicios que, como seres humanos, lleven consigo. Además, la explicitación de las razones que llevaron al juez a tomar una determinada decisión facilita la revisión de las sentencias y otorga legitimidad/credibilidad a las resoluciones judiciales.

Sin ser categórico, el autor pondera que en ciertas circunstancias la fundamentación no se verá afectada por la opacidad algorítmica, aunque la transparencia siempre colaborará en el proceso de revisión de una sentencia, por lo que en caso de que sea técnicamente posible, lo conveniente es tender a buscar la transparencia de este tipo de sistemas.

Fin de la pena, debido proceso y equidad

El fin de la pena

La primera propuesta de desarrollar un sistema digital para asistir a los jueces a la hora de dictar sentencia fue realizada en 1971 por John Hogarth,[14] quien sugería el diseño de un algoritmo basado en la acumulación de una base de datos de fallos y sus consecuentes resultados para analizar las probabilidades de posterior reincidencia. Esta idea, si bien era disruptiva, ya presentaba un anclaje muy claro en una teoría de la pena única.

Es que, la propia idea de orientar los esfuerzos de un sistema informático –que será el encargado de emitir sentencia– a anticiparse a la posible comisión de nuevos delitos futuros por parte del justiciable, habla de una decisión de política criminal de definir el sentido de la pena como de prevención especial. Esto, sea o no adecuado a los intereses de la sociedad en general, omite contemplar una función restaurativa, retribucionista o de cualquier otra índole.

14. Hogarth, John, *Sentencing as a Human Process*, Toronto, University of Toronto Press, 1971.

Lo cierto es que en la mayoría de los países el fin de la pena es un terreno en disputa. Tomando como ejemplo mi propia jurisdicción nacional, la Argentina, es inevitable observar que, si bien nuestra Constitución Nacional se inclina por un fin resocializador de la pena, la ley penal de nuestro país y su aplicación exhiben la preeminencia de otros objetivos de corte retribucionista y de prevención. Por ende, sistematizar esta discusión en un algoritmo que pueda fijar criterios estáticos para resolver casos parece, no solo improbable técnicamente, sino también poco factible a nivel social.

Como puede observarse, desde este primer abordaje ya se vislumbra un problema anterior al de la redacción de una sentencia en sí por parte de una máquina y es la necesidad de discutir y establecer los fines de la pena y la función del sistema de justicia penal en general. Sin esta discusión saldada difícilmente puedan fijarse parámetros que permitan elaborar un método matemático de soluciones jurídicas de los tribunales del fuero penal, pues estas carecerán de legitimidad normativa.[15]

Sumado a ello, la falta de un debate saldado en este sentido conduce a que las bases de datos de decisiones judiciales que pueden utilizar los algoritmos para aprender, estén plagadas de contradicciones, dado que algunas se sostienen en argumentos eminentemente retribucionistas mientras que otras los rechazan y enfocan sus decisiones en un objetivo resocializador o de prevención. Estas inconsistencias pueden ser inocuas para muchas decisiones, pero en determinados casos definitivamente producirían respuestas errantes por parte de un asistente artificial.[16]

El debido proceso penal

Otro elemento trascendental para contrastar con estas propuestas es la garantía del debido proceso legal.[17] De acuerdo a nuestros principios constitucionales y a los estándares internacionales, un juicio justo incluye la posibilidad de conocer y contrastar la evidencia que se presenta en un proceso penal, así como el derecho de obtener una sentencia individualizada o dirigida a un individuo puntual.

15. Van Wingerden, Sigrid; Plesnicar, Mojca, "Artificial Intelligence and sentencing. Human against machines", Ryberg, Jesper; Roberts, Julian V. (eds.), *op. cit.*, pp. 232-234.

16. Wisser, Leah, *op. cit.*

17. Riquert, Marcelo, *op. cit.*, pp. 101-131.

Independientemente de que la prueba pueda ser conocida durante el desarrollo del juicio, previo a la sentencia, lo cierto es que una parte de ese derecho tiene que ver con poder conocer qué elementos fueron tomados en consideración para arribar a la sentencia y cómo fueron ponderados y analizados en concreto, y ahí es donde un algoritmo poco transparente puede resultar contraproducente.[18]

Dentro de las consideraciones que abarca el debido proceso, entiendo que debe ponerse especial atención al derecho de defensa en juicio. Cobra sustancial importancia evaluar el lugar que tiene en los cálculos algorítmicos la voz de las personas acusadas y de sus defensas, para evitar construir un procedimiento que las silencie en aras de estandarizar respuestas para casos similares. Pero además, no puede perderse de vista que para los justiciables la sensación de haber sido escuchados otorga legitimidad empírica[19] a la sentencia, por eso la forma en que los fundamentos de esta son comunicados al público es fundamental.

El principio de equidad

Uno de los problemas más trabajados en relación a la utilización de IA en la toma de decisiones judiciales gira alrededor de la dificultad para crear un sistema imparcial y equitativo. La posibilidad de que el algoritmo construya sesgos y/o sistematice prejuicios humanos, profundizándolos, ha llevado a muchos autores a desconfiar de la idoneidad de este tipo de sistemas para intervenir en procesos penales, inclusive en forma de asesores de los jueces.

Como ya he expresado con anterioridad,[20] a mi entender el problema de los sesgos algorítmicos no es otra cosa que la traducción de los prejuicios humanos y desigualdades sociales que subyacen de los datos y sentencias que analiza y pondera la máquina. Esto plantea un desafío particular, siendo el principio de equidad un argumento utilizado tanto a favor de la implementación de herramientas de IA en las sentencias como en contra de ella.

Es que, mientras que establecer parámetros objetivos mediante cálculos del algoritmo pareciera ser una forma de dejar de lado o

18. Wisser, Leah, *op. cit.*

19. Van Wingerden, Sigrid; Plesnicar, Mojca, *op. cit.*, pp. 234-236.

20. Rangugni, María Catalina, *op. cit.*

al menos visibilizar los sesgos humanos, los sistemas de IA también traen consigo el riesgo de agudizar sesgos de manera solapada.

Esto es así, en primer lugar por lo que algunos autores han convenido en llamar como "problema del *input*"[21] o "bucle de retroalimentación perniciosa".[22] Lo cierto es que las máquinas no tienen la capacidad de realizar un razonamiento circunstanciado de la misma manera que lo hace un humano, por lo tanto, basarán sus predicciones o soluciones en "*inputs*" o información extraída del pasado. De esta manera, lo que puede suceder es que sistemas de esta índole reproduzcan patrones negativos de la conducta humana, que se relacionan con una matriz socioeconómica desigual, o que –por cuestiones técnicas– no puedan evitar reunir información sesgada, porque la información global no está disponible.

De esta manera, una respuesta que pueda parecer "neutral" por parte de la IA en realidad hace mella en prejuicios humanos, o incluso crea nuevos en base a la dificultad para reunir cierta información. Lógicamente los modelos de IA que se desarrollan en la actualidad contemplan esta realidad, y esto da lugar a estrategias de diseño o de gobernanza de datos que buscan revertir estos patrones.

Ahora bien, este aspecto merece especial atención, porque en determinadas circunstancias no bastará con anonimizar la información o eliminar los datos concretos que relacionamos directamente con el sesgo (por ejemplo: el género o la etnia de las personas). En determinadas circunstancias puede haber factores vinculados a ese dato, que hagan de "índice proxy" para identificarlo y reproducir el sesgo de todos modos. Para ilustrar este caso, podemos pensar en el siguiente ejemplo: Una empresa desarrolladora puede conformar una base en la que los datos de etnia o color de piel sean eliminados, para evitar que el modelo algorítmico los tenga en cuenta. Sin embargo, puede darse que el algoritmo identifique patrones dentro del contexto demográfico o laboral que reproduzcan –en los hechos– el mismo sesgo, si, por ejemplo existieran ciertas localidades o ciertas tareas asignadas usualmente a esa etnia o color de piel.[23]

21. Schearze, Mathis; Roberts, Julian, *op. cit.*

22. O'Neil, Cathy, *Armas de Destrucción Matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia*, Madrid, Capitán Swing, 2018.

23. Wisser, Leah, *op. cit.*

¿Es posible esquivar las zonas de controversia?

Es claro que la vocación de los proyectos que actualmente se encuentran en desarrollo para asistir a los jueces en la redacción de sentencias buscan encontrar, eventualmente, soluciones para los problemas descritos.

Independientemente de esa tarea, pienso que es válido también preguntarse si incluso con las tecnologías existentes y sus limitaciones no hay funciones específicas para asignarles que esquiven los ejes conflictivos de debate. En este terreno es en el que, considero, el consenso para avanzar hacia la innovación de las herramientas que utiliza la justicia puede ser mayor.

Por lo pronto, nada obsta la posibilidad de darle a modelos algorítmicos de proyección de sentencias una función de "revisora" de las sentencias redactadas por humanos. De esta manera, podría delinearse un sistema que detecte nulidades, violaciones al debido proceso, y patrones de conducta que repitan determinado sesgo discriminatorio en las sentencias penales.

Más allá de facilitar la tarea recursiva y de revisión, autores como Roberts y Schearze[24] señalan que este tipo de herramientas son fundamentales para la confección de lineamientos unificados de las sentencias que contrarresten sesgos indeseados. De esta manera, al contar con estas correcciones, la idea de construir en el futuro una base de datos de sentencias sin los problemas a los que me referí en el capítulo anterior parece un poco menos improbable.

Modelos aplicados

Habiendo planteado, a grandes rasgos, algunos hilos de discusión, me parece sensato poner la mirada en sistemas que ya fueron diseñados, desarrollados y/o puestos en funcionamiento en distintas latitudes, que se esfuerzan por evitar o contrarrestar las áreas más pantanosas de la IA en el ámbito de la justicia penal.

24. Schearze, Mathis; Roberts, Julian, *op. cit.*

Automatización de trámites sencillos alrededor de la sentencia: República Argentina

Históricamente el Poder Judicial de Argentina ha presentado grandes dificultades en la mecanización de datos para la obtención de estadísticas útiles. Los esfuerzos por resolver esta carencia llevaron, en octubre de 2016, a la firma del Convenio Interjurisdiccional de Datos Judiciales Abiertos y a la creación del Sistema de Datos de la Justicia Argentina (SDJA).[25] Si bien se han dado algunos pasos que permiten un incipiente desarrollo de sistemas de IA que trabajan con datos, todavía hay un largo camino por recorrer para tener modelos confiables que garanticen precisión y diversidad en el *input*.

A pesar de esta circunstancia, algunos modelos han prosperado fuera del ámbito penal, y otros incluso se han proyectado en su interior.

En el ámbito Civil, la Comisión de Informática de la Cámara Nacional de Apelaciones aprobó por unanimidad el Proyecto Hodor, que es un sistema de IA que desarrolló el Departamento de Inteligencia Artificial de la Unión de Empleados Judiciales de la Nación, que automatiza despachos sencillos para impulsar el avance de las causas hacia su resolución. De esta forma, se implementó en el Poder Judicial el primer programa que busca rediseñar e incorporar tecnologías que potencien los trámites judiciales.[26]

Por su parte, la Corte Suprema de Justicia de la provincia de Buenos Aires también ha implementado un sistema informático llamado Augusta: Es un Sistema de Gestión Integral en el cual se registran datos de los casos, así como los pasos procesales, las partes o personas intervinientes, la documentación anexa y toda aquella información que contribuya a la gestión del mismo. Su función, en concreto, es asistir en el despacho del organismo con una biblioteca de modelos propias al organismo y/o genéricas. El sistema cuenta con la posibilidad de

25. Bustos Frati, Gonzalo; Gorgone, Bruno, *Preparación del sector judicial para la inteligencia artificial en América Latina, Caso Argentina*, Buenos Aires, Centro de Estudios de Tecnología y Sociedad (CETyS) de la Universidad de San Andrés, 2021.

26. Méstola, Mariano, *Algunas reflexiones sobre la aplicación de herramientas de automatización en el fuero civil de la justicia nacional*, SAIJ-INFOJUS, 2023. También puede consultarse el sitio web del proyecto, disponible en: https://www.proyectohodor.com.ar/ [Enlace verificado el 16/09/2024]

agendar vencimientos e hitos destacables así como también cuenta con la opción de calcular plazos judiciales.[27]

Si bien "Augusta" es una herramienta sencilla, actualmente se está desarrollando una nueva versión más avanzada denominada "Proyecto Experticia" que sugiere, a través de un algoritmo, distintos proyectos de resoluciones de acuerdo a los datos ingresados por los operadores judiciales. Así, podría estandarizar procesos de trámites simples y cotidianos que no requieren mayor complejidad.[28] Este sistema fue creado como complemento del sistema Augusta para la automatización de diferentes procesos dentro del expediente judicial basado en la confección de árboles binarios, los cuales recibirían como parámetros de entrada referencias almacenadas en el sistema Augusta y entregarían como resultado soluciones con los documentos electrónicos correspondientes, permitiendo automatizar el proceso de decisión en función del estado y datos del expediente electrónico.[29]

En el ámbito del Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires se encuentra la experiencia más difundida a nivel nacional que es el modelo PROMETEA, que cuenta con funciones para agilizar el trámite de los expedientes judiciales y proyecta modelos de resoluciones en base a los datos que el operador judicial ingresa y las indicaciones que comanda al *software*.[30] Tal fue el impacto que tuvo esta innovación que la Corte Constitucional de Colombia exportó el sistema para crear PretorIA en ese país, con funciones similares.

Tal como expresé, puede apreciarse que la jurisdicción criminal no tiene institucionalizadas aún a nivel nacional herramientas propias del fuero, al menos respecto de la redacción de sentencias, sin embargo ya cuenta con diversos prototipos, proyectos y desarrollos de sistemas operativos.

---

27. Disponible en: https://www.scba.gov.ar/paginas.asp?id=39889 [Fecha de consulta: 21/03/2024].

28. Bustos Frati, Gonzalo; Gorgone, Bruno, *op. cit.*

29. Convenio marco de colaboración recíproca entre la Suprema Corte de Justicia de la provincia de Buenos Aires y la Universidad Nacional de la Matanza. Disponible en: https://pupilacdn.nyc3.cdn.digitaloceanspaces.com/diariojudicial.public/documentos/000/092/360/000092360.pdf [Fecha de consulta: 21/03/2024].

30. Corvalán, Juan Gustavo, *La primera inteligencia artificial al servicio de la justicia: Prometea*, Buenos Aires, Thomson Reuters, 2017.

Es el caso del Juzgado 10 en lo Penal, Contravencional y de Faltas de la Ciudad Autónoma de Buenos Aires, donde empleadas y funcionarias desarrollaron un modelo de aprendizaje automatizado que arroja predicciones sobre el tipo de violencia de género de cada caso –según las definiciones de la ley– a partir de la transcripción de los dichos de la víctima.[31]

Una experiencia similar se dio en el ámbito de la justicia criminal federal, en el que un conjunto de funcionarios, junto con docentes de la Universidad de Buenos Aires y miembros del laboratorio de IA de la misma Universidad desarrollaron un modelo para asistir las decisiones judiciales sobre la libertad de las personas en el proceso penal, con el fin de realizar una exploración académica.[32]

Asistente para la mensuración de la pena: Federación de Rusia

Como se refleja en el reporte de la Asociación Internacional de Derecho Penal, en adelante AIDP,[33] Rusia se encuentra debatiendo la viabilidad de implementar un sistema electrónico para la determinación de las penas en los veredictos judiciales. La premisa subyacente del *software* consiste en establecer una pena proporcionada al riesgo planteado por el autor del delito y al grado de culpabilidad asociado.

Este sistema opera a partir de una matriz de decisión que se compone dentro de un marco normativo, en el cual las circunstancias que mitigarían o agravarían el castigo se valoran mediante asignación de puntos positivos o negativos. Estos valores son definidos en base al principio de razonabilidad. En esencia, el juez identifica los elemen-

31. Según la publicación de fecha 15 de septiembre de 2021 en el portal institucional del Departamento de Información Judicial, dependiente de la Secretaría Ejecutiva del Consejo de la Magistratura de la Ciudad Autónoma de Buenos Aires. Disponible en: https://ijudicial.gob.ar/2021/pablo-casas-presento-historias-de-innovacion-de-la-gestion-judicial-centrada-en-las-personas/ [Fecha de consulta: 21/03/2024].

32. Acselrad, Flora Sofía; Nager, Horacio Santiago; Part, Daniela Romina; Reynoso, Carlos Alejandro; Risetti Delión, Vanesa Maura; Soto, Verónica Andrea; Tatian, Rosario, "Inteligencia Artificial y las decisiones sobre la libertad de las personas en el proceso penal", en Corvalán, Juan Gustavo, *Tratado de Inteligencia Artificial y Derecho*, Buenos Aires, Thomson Reuters - La Ley, 2021, T. II., pp. 383-437.

33. Gubko, Vladislav; Novogonskaya, Margarita; Stepanov, Pavel; Yundina, Maria, *AI and Administration of Justice in Russia*. Disponible en: https://www.penal.org/sites/default/files/files/A-07-23.pdf [Fecha de consulta: 21/03/2024].

tos pertinentes, es decir, las circunstancias atenuantes y agravantes del caso en cuestión, ingresándolos al sistema como "*input*". Posteriormente, el programa realiza el cálculo previamente mencionado, involucrando la suma y resta de las circunstancias que alivian o empeoran la situación, y tomando en consideración elementos como el modo de comisión del delito, las condiciones personales del acusado, la naturaleza del delito imputado, entre otros factores relevantes. A partir de esta evaluación, se establece la magnitud de la pena a imponer y, de ser apropiado según las circunstancias particulares, se definen medidas coercitivas adicionales.

Advertencias de desviación: República Popular China

Otro modelo de IA que suscita interés en el ámbito de estudio es aquel que ha sido implementado a modo de prueba en ciertas localidades de China, según se reporta en el informe de la AIDP.[34] Este *software* incorpora un componente conocido como el módulo de "advertencia temprana de desviación". Apoyándose en los beneficios de la investigación científica, la "plataforma de anticipación de desviaciones en juicios por casos similares", desarrollada en colaboración con la Base de Investigación de Datos Judiciales del Pueblo establecida por el Tribunal Popular Supremo, elabora un algoritmo de sentencia mediante el análisis de numerosos documentos legales. Su función principal es proporcionar una alerta temprana de aquellos casos que exhiben desviaciones significativas, con el objetivo de respaldar técnicamente la estandarización de los juicios.

Para ser más concreta, una vez que el juez determina el veredicto y finaliza la redacción del documento de sentencia, el sistema captura automáticamente dicho documento para someterlo a un análisis mediante IA. Los casos que presentan marcadas discrepancias con otros de características similares reciben una notificación de manera automática. Las causas de estas notorias desviaciones se explican a los jueces a través de tecnología de visualización de datos judiciales, o bien, mediante el análisis de la distribución de casos similares y el grado de desviación en los resultados del juicio.

34. Wang, Haiyang, *AI and administration of justice in China*. Disponible en: https://www.penal.org/sites/default/files/files/A-12-2023.pdf [Fecha de consulta: 21/03/2024].

Asistente de predicción del riesgo creado por la libertad del justiciable: Estados Unidos de América

En los últimos años, tanto desarrolladores del sector privado como de oficinas públicas de Estados Unidos han trabajado sobre distintos modelos para asistir a la justicia penal para decidir sobre la libertad de las personas sometidas a procesos o condenadas.

El desarrollo de VPRAI, PRAXIS, PSA, COMPAS y PATTERN, entre otros, ha permitido a los Tribunales valerse de modelos estadísticos para sugerir el riesgo de reincidencia que presenta una persona condenada o, en algunos casos, determinar el peligro de fuga y/o entorpecimiento del proceso de una persona que está siendo sometida a juicio.[35]

A pesar de que el objetivo principal de estos modelos de IA es identificar personas que no representan un peligro real para la sociedad en general o para el proceso penal en particular, y otorgarles la libertad, a efectos de descomprimir las cárceles, su aplicación ha generado enorme controversia. Esto se debe a que el sistema no está exento de sesgos, que perjudican a algunos sectores particularmente vulnerables de la comunidad, además de que en determinados aspectos se ha cuestionado la transparencia con que opera, en los términos en que me he referido en el capítulo respectivo.

Revisión de recursos: República Federativa de Brasil

Por último, considero que es pertinente traer el caso del proyecto "Sócrates",[36] diseñado para operar en el marco de la justicia de Brasil. El sistema fue desarrollado para seleccionar y clasificar los casos y recursos presentados ante el Tribunal Superior de Justicia, realizando un análisis semántico de los documentos procesales con el fin de facilitar la revisión de decisiones judiciales tomadas en instancias anteriores, identificando casos con materias similares e investigando decisiones judiciales que puedan servir de antecedente para el proceso que se examina.

35. Respecto de estos modelos de IA, he desarrollado extensamente su modo de funcionamiento, las críticas y la evolución en su aplicación en Rangugni, María Catalina, *op. cit.*

36. Informe de gestión del Superior Tribunal de Justicia de Brasil 2018-2019. Disponible en: stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relatório%20de%20gestão.pdf [Fecha de consulta: 21/03/2024].

La idea principal de esta herramienta es permitir la revisión de sentencias apeladas, efectuando un análisis que permita elevar un informe en relación a si el caso puede ser juzgado bajo el sistema de recursos especiales que versan sobre la misma materia de derecho (recursos reiterativos), si la legislación aplicada en las decisiones es adecuada e incluso sugiriendo soluciones utilizadas en procesos similares.

Conclusiones

Establecer un sistema automatizado para la redacción de sentencias que sea transparente, equitativo, confiable y que recoja la voz de las partes parece hoy una tarea difícil de proyectar. Sin embargo, esto no significa descartar de lleno la utilización de estas herramientas.

Especialmente en el marco de la justicia criminal, a la hora de dictar una sentencia, el ejercicio del poder del Estado sobre los ciudadanos está en su punto más alto y requiere de altos niveles de consenso y legitimación para operar. En otras palabras, debe valerse de justificaciones sólidas para que la –siempre desproporcionada– respuesta coercitiva que ejerce sobre individuos sea validada por el entorno.

Por esta razón, resulta atractiva la retórica de la optimización cuantificable que se obtiene de las tecnologías de IA para tomar decisiones en otros campos. Sin embargo, al tratarse de una ciencia social dinámica y orientada a regular las relaciones humanas en sociedad este trasvasamiento se torna algo más complejo, por los problemas que fueron desarrollados en el presente trabajo.

No tiene sentido analizar las nuevas tecnologías aisladas de su contexto, por eso es importante no perder de vista el propósito al que sirven. La idea de reemplazar decisiones humanas en el ámbito judicial por cálculos algorítmicos resulta hoy inerte frente a la infinidad de problemas e injusticias que plantearía. Pero la inteligencia puede proporcionar una herramienta de análisis y de procesamiento de datos para agilizar y mejorar decisiones humanas, por lo que no debe descartarse de lleno su implementación. Herramientas como las analizadas, que proporcionan información extra, análisis *"ex post"* de las sentencias para identificar sesgos, patrones disvaliosos, decisiones injustas; y para unificar criterios en cuanto a la justificación de la pena y en cuanto a la manera de

aplicar la ley pueden transformar de manera radical y positiva el funcionamiento de los sistemas de administración de justicia.

Por supuesto que se trata de un terreno delicado, que puede habilitar la utilización abusiva y distorsionada de herramientas de inteligencia artificial que avasalle derechos individuales de los ciudadanos. Así, la importancia de establecer parámetros regulatorios para poner límites a ese tipo de usos se torna urgente. En palabras de Yuval Noah Harari, la regulación ética de la transformación tecnológica y digital en un mundo interconectado como el que vivimos debe ser global.[37]

37. Harari, Yuval Noah, *Homo Deus: A Brief History of Tomorrow*, Londres, Harvill Secker, 2016.

Participaciones internacionales

Desarrollo de un Sistema de Auditoría de Defensa en base a *Big Data* y herramientas de Inteligencia Artificial para audiencias de control de detención

Gonzalo Eugenio Rodríguez Herbach*

Desarrollo

Este proyecto, llevado a cabo entre los meses de junio y diciembre de 2020, tuvo por objetivo general concebir, diseñar y desarrollar, a partir del análisis de los datos sobre casos de primeras audiencias por control de detención, una solución informática que predice comportamientos en:

a. La posibilidad de prisión preventiva,
b. Los plazos de investigación,
c. La posibilidad de alegación de ilegalidad,
d. La entrega de argumentaciones jurídicas más recurrentes respecto de un grupo de delitos perseguidos.

Dicha solución genera una herramienta de auditoría en línea que puede comparar los comportamientos predichos con las actuaciones y gestiones reales, con objeto de generar información para focalizar mejoramientos progresivos en la prestación del servicio.

La solución informática considera modelos de predicción que procesan gran cantidad de información de nuestro sistema y otros para entregar sus cálculos de probabilidades; asimismo, el sistema de manera permanente puede "reentrenarse" de acuerdo a la información que se vaya generando.

Por otra parte, cuenta con una serie de microservicios que consultan nuestro sistema de defensa, exponiéndole al defensor de manera resumida y ágil (en la pantalla telefónica u otro dispositivo) la

* Abogado Universidad de Chile. Máster en Derecho Penal Universidad de Sevilla.

información principal sobre causas vigentes y terminadas, y entregando argumentaciones jurídicas valiosas para la discusión de medidas cautelares, las alegaciones de ilegalidad de la detención, y argumentaciones de derechos humanos sobre estos tópicos, así como señalamientos de si se trata de un imputado con historial de enajenación mental, de migrantes o de adolescentes.

Toda la información permite "adelantar" el conocimiento del imputado y su causa, para que el defensor pueda optimizar sus tiempos de atención, considerando lo breves y acotados que estos resultan, además de proporcionarle datos valiosos para el desarrollo de la entrevista y la audiencia.

Este proyecto contó con el trabajo del equipo de la Defensoría Penal Pública de Puente Alto de Santiago, con la valiosa colaboración de defensores locales, licitados y asistentes, quienes desarrollaron un trabajo de alto valor técnico y operativo, primero en la generación de información y lineamientos para el trabajo de audios, y luego, en la concepción, diseño y ejecución del "Asistente Virtual" para primeras audiencias.

Consideró una fase piloto, iniciada el 10 de noviembre de 2020, en la Defensoría ya aludida, la que generó información y alcances que, de acuerdo a su pertinencia y a la capacidad del proyecto para absorberlo, fueron revisados y ajustados.

El proyecto contempló acciones de capacitación tanto a defensores, asistentes y equipos directivos, todas instancias desarrolladas de modo remoto. A su vez, y luego del proceso de marcha blanca llevado a cabo hacia fines de noviembre y efectuados los ajustes a la aplicación, se generó una fase piloto en 8 defensorías locales del país (Coquimbo, San Antonio, San Joaquín, La Florida, Talagante, Rengo, Linares, Coyhaique). Para ello, se contó con el apoyo y la guía del equipo de defensores y asistentes de Puente Alto, liderado por su defensora local, Ximena Silva.

Se diseñó una estrategia de Multiplicación, en la que cada defensoría que manejaba el "Asistente Virtual" fue capacitando a nuevas defensorías de su región y, de esa manera, considerar, durante el año 2021, que el "asistente virtual" opere en todo el territorio nacional.

Debe señalarse, por último, que la audiencia de control de detención constituye, en muchas jurisdicciones, y particularmente en Puente Alto, la causa de ingreso prevalente a los servicios de la Defensoría Penal Pública; ello está unido a lo crítico que resulta esta audiencia

respecto de las garantías de los defendidos que arriesgan incluso su libertad frente a las pretensiones punitivas; llevadas a cabo, además, en condiciones limitadas en tiempo y recursos para hacerse cargo de la defensa. Ello llevó a la Defensoría Penal Pública a generar este proyecto, que fue perfeccionado durante 2021 e implementado nacionalmente. La empresa de desarrollo del proyecto fue Price Waterhouse Cooper.

Estado Actual del "Asistente Virtual"

En la actualidad el "Asistente Virtual" no se encuentra operativo. La razón principal se debe a la imposibilidad presupuestaria de contar con servicios externos de manejo de la data que soporta el asistente. En efecto, el diseño del proyecto requiere de "espacio suficiente" en discos que permitan el funcionamiento de los algoritmos con su respectivos reentrenamientos.

Al no disponer de este recurso al interior de la Defensoría Penal Pública, la única posibilidad es contratar este servicio a terceros externos. Sin embargo, dada la contingencia de todo este período se consideraron los recursos públicos y se priorizaron en la cobertura las exigencias de la pandemia, por lo que la Defensoría Penal Pública no contó con medios suficientes para su contratación, lo que implicó detener el funcionamiento transitorio del "Asistente".

Por otra parte, durante el tiempo en que el "Asistente" estuvo en uso, la visión de los defensores y defensoras en general se centró en la certeza o no certeza de las cifras que mostraba, en cuanto a las probabilidades que el *software* indicara en torno a la "prisión preventiva", "el plazo de investigación" y la "posibilidad de éxito en la alegación de ilegalidad".

En general, la utilidad del asistente se centró en la información sobre los imputados/as proporcionadas por la aplicación (causas y condenas anteriores, audiencias del artículo 458 del CPP, sobre enajenados mentales, etc.), más que por las cifras proyectadas, circunstancia que aconseja efectuar algunos ajustes en el diseño del "Asistente Virtual" antes de volver a implementarlo; trabajo que ya se había abordado antes de la desactivación.

Anexo

Fuentes de información

1. Oficio Institucional No. 386, de fecha 21 de julio de 2021, del Defensor Nacional de la Defensoría Penal Pública de Chile.
2. Oficio Institucional No. 156, de fecha 21 de marzo de 2023, del Defensor Nacional de la Defensoría Penal Pública de Chile.

Policía y justicia predictiva en España: análisis actual y reflexión crítica*

Jordi Gimeno Beviá**

Policía predictiva

En España no existe una definición uniforme y unívoca del concepto de policía predictiva, sino que la doctrina y práctica forense se sirven de definiciones internacionalmente desarrolladas. Una de los más utilizadas es la recogida por la Organización para la Seguridad y Cooperación en Europa (OSCE) en 2017, la cual define este fenómeno como "la recopilación y evaluación sistemática de datos e información, a través de un proceso analítico definido, que los convierte en productos analíticos estratégicos y operativos que sirven de base para un proceso decisorio mejorado, fundamentado y documentado".[1]

Una aproximación de la doctrina nacional a la conceptualización del fenómeno lo ofrece Miró Llinares, quien incluye la "policía predictiva" dentro de la Inteligencia Artificial Policial y la define como la "aplicación de técnicas cuantitativas para identificar objetivos de interés policial con el propósito de reducir el riesgo delictivo mediante la prevención de delitos futuros o la resolución de delitos pasados".[2]

---

* Este trabajo se corresponde sustancialmente con la publicación de este autor titulada "Instrumentos actuales de policía y justicia predictiva en el proceso penal español: análisis crítico y reflexiones *de lege ferenda* ante aplicaciones futuras" publicada en el número especial de la revista *Estudios Penales y Criminológicos* titulado "Inteligencia Artificial y Sistema Penal" (2023).

** Profesor Titular Derecho Procesal. Véase: http://e-spacio.uned.es/fez/eserv/bibliuned:DptoDPROC-FDER-Articulos-Jgimeno-0001/Gimeno_Bevia_Jordi_Policia_y_justicia.pdf. Todo ello deriva de mi investigación realizada como relator nacional del grupo español de la AIDP (Sección III).

1. Guía sobre actividad policial basada en la inteligencia, OSCE, 2017, p. 6. Disponible en el siguiente enlace https://www.osce.org/files/f/documents/6/4/455536.pdf [fecha de consulta: 18/03/2024].

2. Miró Llinares, Fernando, "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", en *Revista Penal de Derecho y Criminología*, Madrid, Fundación Universitaria Behavior & Law, julio 2018, pp. 87-130.

Se trata, por consiguiente, de un fenómeno relativamente novedoso en nuestro país pues España tiene una experiencia muy limitada –quizás por la ausencia de una regulación específica– en la aplicación de IA para la predicción policial. Y ello porque, tanto en inteligencia artificial como en otras materias, mantiene un enfoque prudente y parece que actuará en línea con los Estados miembro de la Unión Europea y actualmente, como es sabido, el Parlamento Europeo se muestra harto cauteloso a la hora de implementar estos sistemas de vigilancia masiva.[3]

Así las cosas, internamente, el debate no deja de resultar actual pues en nuestro parlamento se presentó por el grupo parlamentario Unidas Podemos una Propuesta no de Ley (PNL) sobre el uso de la Inteligencia Artificial (IA) en las labores de vigilancia y uso de datos personales de la ciudadanía por parte de las Fuerzas y Cuerpos de Seguridad del Estado, además de proponer la creación de una agencia de control de algoritmos para garantizar su transparencia. Y ello porque, desgraciadamente, muchas de las "soluciones" –tal y como suelen definirse a estos sistemas de IA– pretenden ser implementadas por empresas privadas en instituciones públicas, principalmente locales y/o provinciales, mediante contratos públicos con una concurrencia muy limitada pues son muy pocas las empresas nacionales especializadas en el uso de las nuevas tecnologías, lo cual puede ofrecer una situación de incertidumbre en términos de seguridad jurídica.

Más allá de lo anterior, existen instrumentos que las Fuerzas y Cuerpos de Seguridad del Estado (de ahora en adelante, FCSE) emplean habitualmente a la hora de llevar a cabo sus pesquisas basados en el uso de inteligencia artificial.[4]

Seguidamente, expondremos los distintos instrumentos que, para una mejor clasificación, agruparemos en dos bloques. En primer lugar destacaremos los sistemas utilizados actualmente y de un modo generalizado por las Fuerzas y Cuerpos de Seguridad del Estado (FCSE), que son Veripol y Viogen; y, en segundo, aquellas iniciativas

3. Resolución del Parlamento Europeo alertando de los riesgos para nuestro sistema de garantías y libertades de estas tecnologías. Disponible en: https://www.europarl.europa.eu/news/es/press-room/20210930IPR13925/uso-policial-de-la-inteligencia-artificial-el-pe-contra-la-vigilancia-masiva [fecha de consulta: 18/03/2024].

4. Álvarez, Jose Luis, "Policía predictiva en España. Aplicación y retos futuros", en *Behaviour & Law Journal*, 2020, pp. 26-41.

que bien no gozan de un uso tan generalizado, han sido iniciadas pero posteriormente abandonadas o se ha tratado únicamente de experiencias piloto pero de las que conviene advertir su incidencia no solo en los derechos fundamentales de los afectados sino en la propia convivencia ciudadana.

Sistemas utilizados actualmente por las FCSE

VioGen

Quizás en España la herramienta más reconocida en aplicación de IA sea el programa VioGen, destinado a la prevención frente a la violencia de género. Aunque fue creado en el año 2007, el área de Violencia de Género de la Secretaría de Estado de Seguridad (Ministerio del Interior) incorporó inteligencia artificial en el año 2020 a través de la plataforma analítica de la empresa de *software* SAS Iberia.[5]

Su funcionamiento es relativamente sencillo: la aplicación VioGen posibilita la cuantificación los niveles de riesgo de agresión en casos de violencia de género, permitiendo su predicción y consiguiente protección, mediante distintos niveles previamente definidos de las potenciales víctimas. Concretamente, VioGen parte de un primer cuestionario denominado VPR (Valoración Policial del Riesgo) y de un segundo, VPER (Valoración Policial de Evolución del Riesgo), que se realiza de forma general un año después del primero, cuando normalmente ya se ha celebrado el juicio y sea cual sea la sentencia. De esa evaluación se extraen cinco niveles de riesgo: no apreciado, bajo, medio, alto y extremo. Cada uno de ellos lleva aparejado un protocolo de actuación del que la denunciante o víctima es partícipe, así como todas las administraciones con competencias en la materia y el juzgado que instruye o ha juzgado la causa.

Resulta más complejo, sin embargo, medir o cuantificar el éxito de cualquier instrumento cuando se trata de evaluar su impacto en una lacra como lo es la violencia de género. No obstante lo anterior, las autoridades consideran que el porcentaje de fiabilidad de VioGen es relativamente satisfactorio: desde la puesta en funcionamiento de

5. Disponible en: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2020/151220-inteligencia.aspx

esta herramienta hace ya una década la reincidencia de las agresiones ha disminuido un 25% según los últimos datos. Y si de forma genérica la reincidencia en otros países del entorno alcanza el 35%, en España se sitúa ya en el 15%.[6] Los datos se encuentran actualizados en la página web del Ministerio del Interior, lo que permite realizar un análisis sobre la fiabilidad y el éxito de la herramienta desde una perspectiva estadística.[7]

Pero, más allá de la evaluación constante por parte del Ministerio del Interior, VioGen también está siendo evaluado externamente. Sirva como ejemplo la evaluación que está realizando la organización sin ánimo de lucro Eticas Foundation, pero no a solicitud del Ministerio, sino de manera autónoma y con la información publicada así como con la ayuda de fundaciones externas. De hecho, expone las dificultades que tiene para llevarla a cabo pues, según la organización esa falta de transparencia y de explicabilidad implican que no podamos conocer si VioGen tiende a estimar un riesgo demasiado alto o demasiado bajo en ciertos tipos de casos, como podría ser cuando las denunciantes pertenecen a un particular grupo social, como por ejemplo inmigrantes que hablen español (o valenciano o gallego) de un modo diferente a como suelen expresarse quienes hablan el idioma desde siempre.[8]

Y más allá de la referida problemática vinculada a la transparencia, debiera resultar conveniente preguntarse qué es lo que realmente le interesa a la policía judicial, fiscalía y judicatura conforme a la experiencia acumulada durante estos años, pues ello permitiría incorporar datos

6. Entrevista realizada por el periódico *La Vanguardia* a uno de los creadores y Jefe de área de VioGen, Juan José López Ossorio en el año 2017. Disponible en: https://www.lavanguardia.com/tecnologia/20190519/462147339117/viogen-violencia-de-genero-violencia-machista-inteligencia-artificial-algoritmos.html [fecha de consulta: 18/03/2024]

7. Los datos de 2022 se encuentran disponibles en: https://www.interior.gob.es/opencms/es/servicios-al-ciudadano/violencia-contra-la-mujer/estadisticas-sistema-viogen/ [fecha de consulta: 18/03/2024].

8. Disponible en: https://eticasfoundation.org/es/viogen-un-algoritmo-para-predecir-el-riesgo-de-reincidencia-en-casos-de-violencia-de-genero/ [fecha de consulta: 18/03/2024].

que contribuyan a una adecuada tutela de la víctima sin que se produzca vulneración alguna en las garantías constitucionales y procesales.[9]

Por último, en cuanto a su recepción jurisprudencial, los pronunciamientos por parte de las autoridades judiciales son, si cabe, realmente escasos y poco relevantes. Sin embargo, cabría destacar la STS, Sala Quinta, 73/2020, de 28 de octubre, en la que se confirmó la condena a un Guardia Civil que se negó a darse de alta en el sistema VioGen, infringiendo la instrucción de la Secretaría de Estado de Seguridad que obligan a la utilización de los formularios de esta herramienta en la primera evaluación del riesgo de violencia (VPR). Del mismo modo, existe otro pronunciamiento de la Audiencia Nacional, concretamente la Sala de lo Contencioso-Administrativo, en la Sentencia de 30 de septiembre de 2020 en el que se condena a la Administración patrimonialmente tras el asesinato de una mujer por su pareja y no haber comprobado los agentes importantes extremos que hubieran permitido inferir el riesgo existente de reincidencia por parte del agresor hacia la víctima. Como se puede observar, no se ha entrado a valorar cuestiones referentes al "fondo" de los sistemas de predicción policial sino que las dos sentencias expuestas dejan patente un problema de base en la utilización de estos sistemas: la falta de formación (e incluso, en algún caso, desinterés), por parte de los miembros de las FCSE encargados de su uso. Por tanto, más allá de las mejoras técnicas en la aplicación de la IA, debiera perseverarse en la formación de sus usuarios, e incluso, por qué no, en una mayor familiarización con la misma por parte de los operadores jurídicos.

Veripol

El sistema Veripol, puesto en marcha en 2018, está enfocado a prevenir las denuncias falsas, que encuentran su encaje, como es sabido, en el artículo 457 CP. Además de ser la primera herramienta de este tipo en el mundo, cuenta con una precisión de más del 90% y estima la probabilidad de que una denuncia por robo con violencia e intimidación o tirón sea falsa, disuadiendo, entre otras actuaciones –nunca mejor dicho– a denunciantes "espurios", por ejemplo, aquellos que se

9. Llorente Sánchez, "*Big Data*, inteligencia artificial y violencia de género", en *Diario La Ley*, 2021.

inventan el robo de un móvil al único efecto de cobrar el seguro previamente contratado.

Para ello, la herramienta se alimenta de una gran cantidad de datos (*big data*) y determina, con base en el contenido de la información suministrada, el porcentaje de posibilidades de falsedad de la denuncia, utilizando técnicas de procesamiento de lenguaje natural (PLN).

Para su puesta en marcha, la aplicación superó distintas pruebas de funcionamiento, nutriéndose de un banco de más de 1000 denuncias por robo con violencia e intimidación y robo con hurto que fueron presentadas en España durante el año 2015. De las referidas denuncias, aproximadamente un 50% eran verdaderas y otro tanto falsas. El modelo, en el que diversos funcionarios trabajaron durante más de dos años, permite apreciar las diferencias que pueden existir entre la narración de denuncias que han resultado verdaderas y falsas, con base en la información suministrada por el denunciante, la morfosintaxis y una amplia cantidad de detalles.

No obstante lo positivo de la herramienta, algunas voces autorizadas dejan al descubierto importantes deficiencias. Así, pues, en palabras de Jaume Palasí

> El lenguaje corporal también importa en la denuncia y aquí no aparece. Este sistema crea tipos ideales. No describe la realidad, sino que, de forma artificial, establece una descripción mecanizada de la realidad. La realidad es más dinámica que solo unas palabras.

Del mismo modo, que algunos califiquen el porcentaje de 91% de acierto como un éxito, es visto por otros, como Baeza Yates pues, en sus propias palabras "Que se equivoque un 9% implica que el sistema acusa erróneamente a nueve de cada 100 personas. Y esto es un conflicto ético muy grave". Asimismo, los expertos en ética echan en falta una normativa específica, tal y como acontece en otros países (Japón, Finlandia, etc.) que ya han afrontado esta realidad.[10]

Pero, es más, desde una perspectiva netamente procesal parece conculcar la posición de la víctima, de quien un agente, aupado por la aplicación, pone en duda su declaración. Implica, por consiguiente,

10. Ambas opiniones pueden leerse en: https://elpais.com/tecnologia/2021-03-08/veripol-el-poligrafo-inteligente-de-la-policia-puesto-en-cuestion-por-expertos-en-etica-de-los-algoritmos.html [fecha de consulta: 21/03/2024]

un intercambio de roles en el que la víctima de un delito pasa automáticamente a la posición de presunto autor de otro, en este caso, del referido artículo 457 CP. Y, a pesar de que la decisión final recae en el agente, huelga decir que en la mayoría de los casos no se apartará de la decisión del programa.

Sistemas iniciados y/o abandonados por las FCSE

En cuanto a los sistemas cuyo uso no se ha generalizado en las FCSE caben destacar principalmente los SIG o sistemas de información geográfica, los cuales suelen emplearse para prevenir la delincuencia en lugares de alto riesgo mediante una suerte de "mapas digitales del delito" y creación de *hot spots* o puntos calientes en donde se concentra una mayor actividad delictiva.

EuroCop PredCrime

En el ámbito de la seguridad ciudadana, desde el año 2011, distintas Administraciones Públicas principalmente entidades Locales –policías locales–, estuvieron planteando la posibilidad de dotarse –se desconoce si llegaron finalmente a implantarla o si, una vez implementada, tuvieron que abandonarla por su incidencia en los derechos fundamentales o por la falta de una regulación suficiente– del *software* EuroCop PredCrime. El *software*, tal y como se define en la web, consiste en

> … el desarrollo experimental de un Sistema Integrado de tratamiento de datos masivos vinculados a delitos y faltas ya cometidos, basado en el uso de modelos matemáticos y algoritmos, que permite la prevención y resolución de un crimen aún no producido.[11]

Se trata de un sistema que integra y trata datos masivos vinculados a delitos, que basa su funcionamiento en un modelo espaciotemporal e información geográfica de mapas de calor mediante modelos y algoritmos matemáticos para la prevención, a través de la predicción, de los delitos que, en el futuro, pudieran cometerse.

11. Disponible en: https://www.eurocop.com/catedra-eurocop/proyectos-en-marcha/eurocop-pred-crime-sistemas-para-la-prediccion-y-prevencion-del-delito/ [fecha de consulta: 21/03/2024]

Por otro lado, *software* como el proporcionado por Eurocop PredCrime, empresa privada, fue contratado por ayuntamientos para la protección de sus municipios, si bien, falta publicidad para conocer el alcance de los contratos, así como su objeto concreto y valorar si la aplicación de técnicas de vigilancia computarizada (tal y como parece que emplean) resulta contraria tanto al criterio de la UE como a la propia Agencia Española de Protección de Datos. De hecho, algunos de los sistemas de EuroCop PredCrime fueron "abandonados temporalmente" (intuimos que por falta de garantías o base legal para su uso) por los Ayuntamientos que lo habían suscrito, tales como el de Rivas Vaciamadrid (Madrid).[12]

El principal problema que avanza la utilización de estos *softwares* radica en la participación privada no ya de la seguridad pública –lo cual acontece habitualmente en no pocos recintos– sino en el trasvase y manejo de datos e información sensible acopiada habitualmente en bases de datos policiales. En efecto, estas herramientas que surgen de un partenariado público-privado pueden acarrear profundos problemas de legalidad, de ahí que, a fecha de la redacción del presente trabajo, no se pueda hablar de un uso generalizado por las FCSE sino más bien al contrario, pues dadas las referidas dudas no se tiene constancia de que sean utilizados actualmente.[13]

12. Disponible en: https://www.rivasciudad.es/noticias/organizacion-municipal/2015/12/10/un-sistema-pionero-en-prevencion-de-delitos/862600041423/ [fecha de consulta: 21/03/2024]

13. Muy críticos se muestran al respecto Ekaitz Cancela y Aitor Jiménez, periodistas de *El Salto* que, tras una profunda investigación, alertan de los riesgos que plantea esta herramienta. Así pues, se plantean los siguientes interrogantes, que reproducimos literalmente: "¿A qué datos comprometidos y privados puede tener acceso una compañía que presta y gestiona la infraestructura digital crítica de las agencias de policía? ¿No tienen los ciudadanos derecho a conocer el interior de estas cajas negras? ¿Queremos que una corporación privada esté en posición de ofrecer 'una solución que cubre la gestión integral de la policía, tanto en el aspecto operacional (automatizando todas sus tareas operativas, administrativas, judiciales, etc., desde cualquier lugar y momento), como en el aspecto táctico y estratégico a fin de lograr la máxima eficacia en la labor policial'?...". El resultado de la información, muy crítica con estos sistemas de policía predictiva adoptados por las policías locales. Disponible en: https://www.elsaltodiario.com/tecnologia/estado-policial-espanol-2.0-empresas-privadas-eurocop-vigilar-ciudadanos [fecha de consulta: 21/03/2024]

Predictive Police Patrolling (P3-DSS)

También se desarrolló en el año 2017 un estudio piloto realizado por el Cuerpo Nacional de Policía (CNP) en el distrito centro de Madrid, titulado *Predictive Police Patrolling* (P3-DSS) que, mediante algoritmos matemáticos, permite predecir delitos, conocer su tipología así como mejorar la eficiencia de los turnos de patrullas policiales. El proyecto fue ideado por el policía y matemático Miguel Camacho, y parte del mismo puede verse en su propia tesis doctoral titulada *Statistical Analysis of Spatio-Temporal Crime Patterns: Optimization of Patrolling Strategies*, defendida en 2016.[14]

Esta aplicación referida a la prevención del delito y la mejora de la eficiencia en el patrullaje, mediante el desarrollo del Sistema de Información Geográfica (SIG) permite a la policía gestionar en un tiempo razonable datos espaciotemporales que ayudan a identificar concentraciones de hechos delictivos permitiendo, por tanto, implementar un "patrullaje predictivo" que dota de una mayor eficiencia la distribución de patrullas en función del riesgo criminal.[15] Para el uso de la herramienta piloto utilizada en el Distrito Centro de la ciudad de Madrid, se recopilaron los registros criminales referentes al delito de robo (105.755 incidentes) entre los años 2008 y 2012. A su vez, se valieron de los Sistemas de Información Geográfica (SIG) del CNP que integran los sucesos delictivos sobre un mapa geográfico de la ciudad, además de la localización de las patrullas de policía.[16]

Sin embargo, estas técnicas de predicción policial han sido cuestionadas porque pueden colisionar con el derecho a la igualdad y a la no discriminación. Efectivamente, quizás una de las primeras cuestiones a debatir sobre la inteligencia artificial para la predicción policial radique en el derecho a la igualdad o no discriminación traducida en

14. Disponible en: https://hera.ugr.es/tesisugr/26134081.pdf [fecha de consulta: 21/03/2024]

15. Álvarez, José Luis, *op. cit.*, p. 30.

16. Jiménez Hernández, M., "El *Big Data* como herramienta de prevención de la delincuencia", Universidad de Alicante, p. 28. Disponible en: https://rua.ua.es/dspace/bitstream/10045/115934/1/EL_BIG_DATA_COMO_HERRAMIENTA_DE_PREVENCION_DE_Jimenez_Hernandez_Miguel_Angel.pdf [fecha de consulta: 21/03/2024]

el riesgo de que esta contenga determinados sesgos en los algoritmos que la conforman.

De hecho, la Carta de Derechos Digitales, más allá de que no tenga fuerza normativa, dispone en su derecho XXV "Derechos ante la inteligencia artificial", concretamente en su apartado 2.a) que "Se deberá garantizar el derecho a la no discriminación cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en inteligencia artificial".[17]

Y esta preocupación también ocupa a la doctrina. Así, pues, Nieva Fenoll alerta sobre el uso de *big data* en la investigación policial porque los datos que aleatoriamente se almacenan de personas, barrios, etc., pese a dicha aleatoriedad, se habrán seleccionado en función de los propios perjuicios del configurador del algoritmo, lo que implica que los resultados no sean neutrales.[18] También, de un modo concluyente, Miró Llinares resume el problema planteado:

> … las herramientas predictivas de las que hablamos no vienen más que a hacer lo que ya se hacía y se hace a día de hoy de manera tradicional y manual y probablemente con los mismos sesgos o más, añadiendo, en algunos casos, una metodología más sistemática o científica pues, –prosigue el autor– […] lo que sabemos hasta el momento nos dice que los algoritmos, que reflejan con precisión nuestro mundo, parecen reflejar también nuestros prejuicios.[19]

Por su parte, Camacho-Collados, González-Álvarez y Santos Hermoso exponen el debate existente aludiendo, de un lado, a que, si bien los defensores de este tipo de prácticas policiales argumentan que estas herramientas ayudan a reducir los sesgos de los investigadores, ya que sustituyen la experiencia personal por un conocimiento basado en el análisis sistemático de todos los casos conocidos y esclarecidos; sus detractores entienden que en tanto estas herramientas se basan en delitos resueltos previamente por unidades policiales, puede haber determinados delitos con una mayor probabilidad de ser investigados y

17. Llorente Sánchez Arjona, Mercedes, *Digitalización de la justicia: prevención, investigación y enjuiciamiento,* Navarra, Thomson Reuters Aranzadi, 2022, p. 179.

18. Nieva Fenoll, Jordi, "Inteligencia artificial y proceso judicial", Madrid, Marcial Pons, 2018, p. 151.

19. Miró Llinares, Fernando, *op. cit.*, p. 126.

solucionados, porque son cometidos por individuos que se consideran peligrosos, o porque tienen lugar en zonas propensas a la delincuencia. Si los datos que se introducen en el modelo presentan algún tipo de sesgo, se podría provocar la estigmatización de individuos o zonas que presenten las características que los algoritmos identifican como indicadores de riesgo. No obstante el debate existente, en línea con los autores anteriormente citados, entienden que el desarrollo de estas prácticas es la evolución lógica de la práctica policial tradicional.[20]

Ahora bien, sin desdeñar los potenciales beneficios que puede aportar la policía predictiva, pues la prevención siempre es preferible a la reacción, la estrategia policial, como acertadamente sugiere Guzmán Fluja, debiera combinarse con acciones políticas, sociales y económicas que pongan a la persona en el centro, absteniéndonos de su consideración –como parece acontecer con las herramientas de *predictive policing*– como delincuente o potencial delincuente.[21]

Justicia predictiva

Una apuesta de futuro ¿a corto plazo?

A la hora de realizar una primera aproximación contextual al fenómeno, cabe señalar que tampoco existe una definición unívoca de justicia predictiva en tanto no se encuentra contemplada en la legislación española. Por consiguiente, la doctrina y la práctica forense se sirven de su acepción literal, directamente traducida del *predictive justice* para referirse a ella.

No obstante lo anterior, algunos autores relevantes como Armenta Deu califican la justicia predictiva como un "término muy amplio" y consideran que "se promueve como herramienta para la eficiencia procesal, combinando una mejora en la calidad de la toma de decisiones y una reducción de la actividad judicial".[22]

Otros autores asocian su significado al de "jurimetría". En efecto, Peralta Gutiérrez indica que "la Jurimetría es la aplicación de la

20. Álvarez, José Luis, *op. cit.*, pp. 28-29.

21. Llorente Sánchez Arjona, Mercedes, *op. cit.*, p. 301.

22. Armenta Deu, María Teresa, *Derivas de la justicia*, Madrid, Marcial Pons, 2021.

inteligencia artificial y *machine learning* a los tradicionales buscadores legales y jurisprudenciales obteniendo nuevas funcionalidades que dan lugar al desarrollo de lo que también se ha denominado 'Justicia predictiva'".[23]

Sin embargo, aunque son términos relacionados, entendemos que justicia predictiva es un fenómeno más amplio que la jurimetría o, si se prefiere, la segunda forma parte de la primera o consiste en una herramienta o técnica de justicia predictiva, si bien pueden existir otras.[24] Entendemos, en resumidas cuentas, que "justicia predictiva", versa sobre IA utilizada en la función jurisdiccional para la toma de decisiones judiciales.

En cuanto al "estado del arte" sobre la justicia predictiva en España, no existe una negativa rotunda hacia el uso de sistemas de inteligencia artificial. No estamos, por consiguiente, en el mismo escenario que países de nuestro entorno, como Francia, que en el artículo 33 de la Ley para la Reforma de la Justicia prohíbe, estableciendo incluso penas de cárcel, que se publique información estadística sobre las decisiones y el patrón que siguen los jueces a la hora de dictar sentencia.

Mas al contrario, actualmente se encuentra en fase de Proyecto una norma que pretende el uso de inteligencia artificial para la modernización de la Administración de Justicia: el Proyecto de Eficiencia Digital. Se trata de una "acción nuclear" dentro del programa de Justicia 2030, impulsado por el Gobierno de España.[25] De hecho, en enero de 2022 se publicó la licitación del Acuerdo Marco de Justicia para la Transformación Digital con más de 125 millones de euros.[26]

23. Peralta Gutierrez, Adolfo, "Diálogos para el futuro judicial XXIII. Jurimetría y justicia predictiva", *Diario La Ley*, Nº 9837, 26/04/2021.

24. Como sostiene Suarez Xavier "la jurimetría, la estadística judicial, procesos de automación y otras formas de smartificación de la justicia pueden integrarse en el concepto de justicia predictiva, pero no configuran el concepto en sí mismo" en su Tesis Doctoral "Gobernanza, inteligencia artificial y justicia predictiva: los retos de la administración de justicia ante la sociedad en red". Disponible en: https://riuma.uma.es/xmlui/bitstream/handle/10630/20979/TD_SUAREZ_XAVIER_Paulo_Ramon.pdf?sequence=1 [Fecha de consulta: 25/03/2024].

25. Bueno De Mata, Federico, *La necesidad de regular la inteligencia artificial y su impacto como tecnología disruptiva en el proceso: de desafío utópico a cuestión de urgente necesidad*, Thomson Reuters Aranzadi, 2021, p. 29.

26. En dicho Acuerdo, el lote 2 está destinado a los proyectos relacionados con la ingeniería de datos; laboratorios de innovación e inteligencia artificial; implantación

Sin embargo, la utilización por los jueces de opciones de jurimetría o técnicas analíticas de datos, si bien resulta factible, no parece que, al menos a corto plazo, vaya a producirse. Tampoco parece que vayan a implementarse herramientas de justicia predictiva "decisorias" del órgano judicial, para la imposición de medidas cautelares, dictado de sentencias, etc. pues la utilización real de la IA mediante técnicas auxiliares como las referidas implica una transformación sustancial de nuestra justicia que requiere reformas de mayor calado que las planteadas en el Proyecto de Eficiencia Digital, dejando a un lado únicamente la *rara avis* que supone RisCanvi utilizado únicamente en el sistema penitenciario catalán que, seguidamente, expondremos.

Sistemas actualmente utilizados en la justicia penal

RisCanvi

Se trata de una herramienta empleada en fase de ejecución de la pena para la predicción del riesgo de violencia a la hora de conceder permisos y determinar la situación del penado. Podríamos clasificarla dentro del ámbito de la justicia predictiva, porque evalúa unos ítems determinados para arrojar el riesgo de una determinada futura actuación pero no aplica inteligencia artificial en sentido estricto.[27] De otro lado, su ámbito de aplicación es limitado ya que sólo se aplica en las cárceles de Cataluña donde, a diferencia de lo que acontece con otras Comunidades Autónomas, las competencias en materia de prisiones fueron transferidas.

El sistema funciona aplicando dos protocolos, en primer lugar, el RisCanvi *screening*, que se aplica a la entrada del penado al centro

---

de soluciones basadas en inteligencia artificial e ingeniería de datos para identificar, extractar y explotar información dentro del ámbito judicial para el establecimiento y mejora de modelos predictivos; así como la definición de soluciones tecnológicas disruptivas que ayuden a la Administración de Justicia a cumplir con sus objetivos. Disponible en: https://www.mjusticia.gob.es/es/ministerio/gabinete-comunicacion/noticias-ministerio/220113-NP-Justicia-publica-la-licitacion-del-Acuerdo-Marco-de-Justicia [fecha de consulta: 21/03/2024].

27. Férez-Mangas, David; Andrés-Pueyo, Antonio, "Eficacia predictiva en la valoración del riesgo del quebrantamiento de permisos penitenciarios", *La Ley Penal*, N° 134, septiembre 2018, p. 8.

penitenciario y que consta de diez ítems distintos y que arroja un riesgo bajo o alto de violencia y el RisCanvi *complet*, que ofrece mayores garantías porque cuenta con 43 variables que arrojan una puntuación de riesgo bajo, medio o alto. Las funcionalidades de la herramienta se han ampliado hasta cuatro: quebrantamiento de condena, la violencia dentro de la prisión, la reincidencia violenta y la violencia autodirigida (suicidio, intento de suicido, autolesiones...).[28] Los resultados de la evaluación son confirmados o modificados por la junta de tratamiento penitenciario (que en muy pocas ocasiones se apartan de ella)[29] y suelen guiar la decisión final del juez a la hora de conceder los permisos penitenciarios, la libertad condicional, la clasificación del penado y la adopción de medidas de supervisión.[30]

Sin embargo, RisCanvi no utiliza estrictamente IA sino sistemas de regresión logística, de ahí que, como indica uno de los responsables de los Sistemas Penitenciarios de Cataluña, la herramienta tenga mucho margen de mejora y que pronto incorporará técnicas de *machine learning* o modalidades más sofisticadas que las actuales,[31] limitadas a la alimentación del algoritmo a través de las valoraciones de los internos que añaden los funcionarios y de las propias bases de datos de las cárceles (días de condena, sentencias, etc.).

En cuanto a la aplicación de la herramienta, su porcentaje de uso alcanza casi el 100% pues sólo en un 3,2% de los casos la evaluación final ha sido modificada por el evaluador. Incluso, se critica que aun cuando el evaluador o equipo técnico del centro penitenciario defienda la decisión de otorgar un permiso penitenciario, si ello acontece en contra de RisCanvi porque la herramienta aprecie un riesgo medio o

28. Ibídem, p. 6.

29. Concretamente un 3,2% del total. Véanse las estadísticas publicadas en el *Diario La Vanguardia*, Disponible en: https://www.lavanguardia.com/vida/20211206/7888727/algoritmo-sirve-denegar-permisos-presos-pese-fallos.html [fecha de consulta: 22/03/2024].

30. Montesinos García, Ana, *Justicia penal predictiva*, Valencia, Tirant lo Blanch, 2022, p. 440.

31. La información sobre el funcionamiento de RisCanvi así como las opiniones de expertos sobre dicha herramienta, se encuentran disponibles en esta noticia con las respectivas entrevistas. Disponible en: https://www.todonoticia.cl/2021/07/11/prisiones-riscanvi-luces-y-sombras-del-algoritmo-que-ayuda-al-juez-en-cataluna-a-decidir-si-mereces-la-condicional-transformacion-digital-tecnologia/ [fecha de consulta: 22/03/2024].

alto de reincidencia, el Fiscal recurra la decisión amparándose únicamente en el resultado del *software*.[32]

Los problemas o defectos que se imputan a esta herramienta radican en la existencia de falsos negativos lo que lleva a parte de la abogacía a alertar del peligro que supone que la decisión sobre la libertad o no de un penado descanse en un algoritmo que, indefectiblemente, siguen aquellos que deben tomar una decisión final.[33] Bien es cierto que la decisión la adopta el juez tras haber sido informado o haber leído el informe realizado por el equipo técnico, que es quien utiliza esta herramienta, pero no lo es menos que los abogados de los presos desconocen tanto el funcionamiento como los parámetros en los que se basa la aplicación y que en tanto sus clientes ya han sido condenados, cuentan con menos posibilidades de rebatir la pertinencia de obtención de un determinado permiso frente a un instrumento científico y teóricamente objetivo. Por otro lado, que pueda existir algún falso negativo y que el penado delinca no debiera ser motivo para replantear una herramienta a través de las cuales, estadísticamente, se conceden más permisos que antes de su existencia.[34]

Sistemas de jurimetría y analítica de datos judiciales

En diciembre de 2021, el CGPJ puso a disposición de los Jueces y Magistrados –también miembros de la Oficina Judicial, como LAJs– la aplicación KENDOJ (*Knowledge Extractor* for CENDOJ) que aplica técnicas de IA y *machine learning* para llevar a cabo dos actuaciones: de un lado, la pseudonimización automática de un documento para el cumplimiento de los estándares del RGPD y, de otro, facilitar el acceso a la información más relevante de cada documento con anterioridad a su

32. Si bien no podemos comprobar este extremo, así lo defiende el Magistrado y Catedrático de Penal Daniel Varona. Disponible en: https://www.lavanguardia.com/vida/20211206/7888727/algoritmo-sirve-denegar-permisos-presos-pese-fallos.html [fecha de consulta: 22/03/2024].

33. En palabras de Marisa Díaz, abogada de derecho penitenciario: "La ley es clara sobre las condiciones para pedir permisos o acceder al tercer grado, y no dice nada del Riscanvi. Utilizarlo para denegar es una vulneración grave de derechos". Disponible en: https://www.lavanguardia.com/vida/20211206/7888727/algoritmo-sirve-denegar-permisos-presos-pese-fallos.html [fecha de consulta: 22/03/2024].

34. Martinez Garay, Lucía, "Errores conceptuales en la estimación del riesgo de reincidencia", en *Revista Española de Investigación Criminológica (REIC)*, Nº 14, 2016.

lectura así como una búsqueda más acertada de legislación y jurisprudencia para el desarrollo de sus funciones.[35]

Pero más allá de este instrumento con tan limitado alcance, lo cierto es que otros operadores jurídicos como son los abogados, ya disponen de herramientas o sistemas de *LegalTech* basados en justicia predictiva para el ejercicio de su función. Se trata de *software* especializado creado por potentes empresas del sector jurídico que, a grandes rasgos, permiten al usuario predecir las posibilidades de éxito de sus actuaciones ante un determinado Tribunal, basado todo ello en *big data* judicial obtenidos del CENDOJ, tras suscribir acuerdos de explotación comercial con el CGPJ.

Sin ánimo de ahondar en todas las aplicaciones jurímétricas o de analítica de datos jurídicos, al efecto de exponer su funcionamiento, sirva como ejemplo Jurimetría, del grupo Wolters Kluwer, seguramente una de las aplicaciones pioneras (año 2017) en la aplicación de la justicia predictiva al Derecho. Se trata de una herramienta que sistematiza y extrae de forma exhaustiva la inteligencia que reside en un conjunto de más de 10 millones de resoluciones judiciales y en toda la estadística judicial procedentes de todas las instancias y órdenes jurisdiccionales de España, a las que se incorporan medio millón de nuevas resoluciones cada año. Jurimetría consta de seis módulos interconectados, cada uno con una finalidad y alcance diferente y complementario: 1) Jurimetría del caso: evalúa los parámetros críticos para el éxito del caso, conociendo la trayectoria del Juez y de los abogados contrarios, con acceso a la jurisprudencia más relevante; 2) Jurimetría del Juez o Magistrado: permite analizar la trayectoria, líneas argumentales y posicionamientos del juez en cuestión; 3) Jurimetría del abogado: análisis global de la contraparte en el proceso, desde todas las perspectivas; 4) Jurimetría de la empresa: permite realizar un análisis de los litigios en los que ha sido parte alguna de las grandes empresas; 5) Jurimetría del Tribunal: permite conocer la actividad de los juzgados y tribunales de España, en aspectos como la duración media de los procesos, la congestión o la probabilidad de recurso; y 6) Jurimetría del Organismo Público: permite examinar los

35. Disponible en: https://confilegal.com/20211218-la-carrera-judicial-contara-desde-el-lunes-con-una-aplicacion-basada-en-la-inteligencia-artificial-y-machine-learning/ [fecha de consulta: 22/03/2024].

procesos judiciales en los que ha sido parte un organismo o entidad pública, a partir de cualquier óptica.[36]

Como hemos avanzado, la utilización de estos sistemas de analítica de datos judiciales –que emplean principalmente los abogados– no está prohibida, como en Francia, sino que, sencillamente, se entiende permitida al no estar regulada. Las grandes editoriales adquieren a través de acuerdos comerciales con el CGPJ todos los datos (sentencias, autos, etc.) de los que nutren a estos productos. Y si bien se utiliza habitualmente desde el año 2017, ha sido este año 2022 el momento en el que el CGPJ se ha quejado o, mejor dicho, ha tomado conciencia, de las implicaciones que pueden tener estos sistemas indicando, literalmente, que

> … se echa en falta la determinación de los criterios de utilización de la jurimetría y de la inteligencia artificial por parte de empresas y particulares a partir de bases de datos, incluso aquella configuradas por este órgano constitucional; el cual habrá de tener la necesaria participación en la determinación de tales criterios de uso.[37]

Resulta, a nuestro juicio, paradigmático que a estas alturas –esto es, cuando los productos llevan años empleándose– el CGPJ, que a la postre tiene acuerdos comerciales de explotación con estas editoriales, muestre su preocupación al respecto.

Pero tal preocupación no es, ni mucho menos, baladí pues también ha sido compartida por la doctrina procesal, si bien refleja distintas posiciones. Armenta Deu alerta de los riesgos que plantea la justicia predictiva, riesgos que califica de "sistémicos" pues implicará que el justiciable deje de acudir –o acuda menos, según sus palabras– a los tribunales habida cuenta los pronunciamientos ya serán conocidos y la justicia predictiva conseguirá que éstos sean cada vez

36. Información extraída de su propia página web. Disponible en: https://jurimetria.laleynext.es/content/Inicio.aspx# [fecha de consulta: 22/03/2024].
También existen otras aplicaciones interesantes como son Tirant Analytics. Disponible en: https://analytics.tirant.com/analytics/estaticas/guiausuario/guia_analytics_web.pdf, vLex Analytics. Disponible en: https://vlex.es/p/spain-court-analytics/ y Neo de Lefebvre. Disponible en: https://lefebvre.es/noticia/nace-neo-la-primera-plataforma-la-gestion-del-conocimiento-juridico-del-mercado-europeo/ [fecha de consulta: 22/03/2024].

37. El resumen del Informe del CGPJ al Anteproyecto de Ley de Eficiencia Digital. Disponible en: https://www.poderjudicial.es/cgpj/es/Poder-Judicial/En-Portada/El-Pleno-del-CGPJ-aprueba-por-unanimidad-el-informe-al-anteproyecto-de-ley-de-Eficiencia-Digital-del-Servicio-Publico-de-Justicia [fecha de consulta: 22/03/2024].

más homogéneos.[38] Por otro lado, Nieva Fenoll señala tras describir algunas herramientas de justicia predictiva utilizadas en otros países que, en la argumentación jurídica

> … la inteligencia artificial hará que la labor de persuasión sea menos ardua, al poderse recopilar con mucha mayor facilidad la información disponible y los argumentos a favor y en contra de las diferentes opciones y, como ya se dijo, no estará condicionada por las emociones o sentimientos, sino que integrará solamente datos objetivos.[39]

También Martín Diz considera prioritario el asentamiento de la IA y su aplicación al derecho procesal pero priorizando las garantías frente a la eficiencia "más aún cuando por su innegable grado de avance tecnológico, y por lo que pudiera servir en futuras décadas como elemento de asistencia a abogados y de predictibilidad".[40] Barona Vilar reflexiona acerca de la jurimetría y considera que "hay que valorarla como lo que es, a saber, un sistema computacional asistencial. No es un modelo sustitutivo de la mente humana"[41] y, en sentido parecido, Borges Blázquez entiende que es factible su uso por la autoridad judicial, tal y como la utilizan los abogados, si bien expone los recelos que pudiera generar en el Poder Judicial, dadas las cautelas que ha tenido en asuntos relacionados el Consejo General del Poder Judicial.[42]

En nuestra opinión, produce cierta desconfianza que grandes empresas del sector jurídico marquen el paso a la Administración de Justicia. No cabe duda que servirse de estas herramientas *Legal Tech* para la elaboración de la estrategia procesal colocará a la parte que las posea en una posición aventajada frente a la que carezca de ellas.[43]

---

38. Armenta Deu, María Teresa, *op. cit.*, p. 246.

39. Nieva Fenoll, Jordi, *op. cit.*, p. 30.

40. Martín Diz, Fernando, "Justicia predictiva: inteligencia artificial y algoritmos aplicados al proceso judicial en materia probatoria", en *El impacto de las tecnologías disruptivas en Derecho procesal*, Navarra, Thomson Reuters Aranzadi, p. 138.

41. Barona Vilar, Silvia, *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice*, Valencia, Tirant lo Blanch, 2021, p. 371.

42. Borges Blázquez, Raquel, *Inteligencia artificial y proceso penal*, Navarra, Thomson Reuters Aranzadi, 2021, p. 164.

43. En un mismo sentido se pronuncia Reifarth Muñoz quien explica que "es posible que la inteligencia artificial produzca desequilibrios en la construcción de una línea de defensa eficaz. La preparación de un asunto con sistemas legaltech es mucho más

Y en contra de lo que pudiera pensarse, no sólo se aplica en el orden jurisdiccional civil, sino también en el penal, donde podrá ocurrir que el investigado que carezca de recursos y se acoja a la justicia gratuita se enfrente no solo al Ministerio Fiscal –que quizá a título particular pueda emplear esta herramienta– sino también al acusador particular provisto de estas tecnologías. Y en un contexto en el que la jurisprudencia cada vez tiene mayor relevancia en el sistema judicial, corremos el riesgo de avanzar hacia un control, clasificación y comercialización del *big data* judicial por parte de las grandes tecnológicas, conculcando el principio de igualdad de armas y relegando a quienes carecen de recursos a un acceso básico –y no *premium*– a la jurisprudencia.[44]

El encaje de la IA en nuestra justicia penal

Una vez expuestos algunas herramientas de policía y justicia predictiva que tímidamente emplean inteligencia artificial, cabe plantearse si una IA más intensa y con un uso más generalizado en las distintas fases del proceso tiene o puede tener encaje en nuestra justicia penal. Para ello, deviene imprescindible, entre otras premisas básicas, partir de las siguientes.

a. Respeto al artículo 117.3 CE y al principio de exclusividad jurisdiccional: el debate no se encuentra tanto en limitar el acceso al juez humano en algunos casos sino en que detrás de la justicia, esto es, de la toma de decisiones, debe haber un juez humano. Y ello no es por capricho, sino porque en el artículo 117.3 de la

sencilla, ya que permite procesar datos, comparativas y probabilidades de éxito o fracaso de un determinado asunto con extraordinaria rapidez. En términos generales, la utilización de estos instrumentos es beneficiosa pero puede generar disfunciones cuando solo una de las partes tiene acceso a ellos", en su texto "El uso de la inteligencia artificial en el proceso judicial y los derechos fundamentales".

44. Salvando las distancias, el problema referido nos recuerda a lo acontecido en los Estados Unidos donde un joven activista, Aaron Schwartz, trató de "liberar" la jurisprudencia de los tribunales federales de EEUU, a la que se accedía previo pago, descargando la jurisprudencia de PACER para alojarla en la base abierta (*open access*) de RECAP. Desgraciadamente, este joven acabó suicidándose pues, entre otros problemas, se enfrentaba a décadas de prisión por las descargas realizadas. Disponible en: https://arstechnica.com/tech-policy/2013/01/internet-pioneer-and-information-activist-takes-his-own-life/ [fecha de consulta: 22/03/2024].

Constitución española se reconoce el principio de exclusividad jurisdiccional por el que "El ejercicio de la potestad jurisdiccional en todo tipo de procesos, juzgando y haciendo ejecutar lo juzgado, corresponde exclusivamente a los Juzgados y Tribunales determinados por las leyes, según las normas de competencia y procedimiento que las mismas establezcan". Del artículo 117.3 CE, indica Borges Blazquez que podemos extraer dos conclusiones: "... la primera, nuestra constitución excluye a otros sujetos o sistemas de la capacidad de juzgar. La segunda, el artículo dice quién debe ejercer la función: jueces y tribunales. Pero no especifica cómo ni mediante qué herramientas. Por tanto, el uso de sistemas de IA de manera complementaria, actuando como apoyo a la decisión que debe tomar el juzgador y nunca sustituyendo su razonamiento tendrían encaje en nuestro sistema".[45]
En esa misma línea se pronuncia Montesinos García quien concluye que "Damos la bienvenida a todo lo que coadyuve a objetivar determinadas decisiones que deben adoptar los jueces. Pero tenemos que partir de la premisa que estamos ante sistemas asistenciales, esto es, ante herramientas de colaboración. La decisión del juez no puede descansar en exclusiva sobre un algoritmo. El resultado que proporcione el sistema predictivo, en el caso de ser considerado por parte del juez, solo podrá hacerlo como un elemento más, que deberá en todo caso ser corroborado por otros elementos del juicio. De modo que concluimos este trabajo afirmando con rotundidad que la función predictiva, en caso de integrarse en la justicia penal debe ser únicamente de apoyo o asesoramiento en la toma de decisiones judiciales, pero nunca podrá alcanzar fuerza decisoria ni ser vinculante para los jueces".[46] Por consiguiente, la inteligencia artificial aplicada a la justicia penal debe ser asistencial, nunca sustitutoria de la función jurisdiccional.[47]

45. Borges Blázquez, Raquel, *Inteligencia artificial y proceso penal*, *op. cit.* p. 197.

46. Montesinos García, Ana, *Justicia poliédrica en tiempos de mudanza: Justicia penal predictiva*, Valencia, Editorial Tirant lo Blanch, 2022, p. 449.

47. Marchena Gómez, Manuel, "Inteligencia artificial y jurisdicción penal". Discurso de ingreso en la RAD, 2022, pp. 38-39. Sobre ello se pronunció con contundencia el Magistrado quien subraya que "No puedo identificarme, sin embargo, con esta línea de

b. La IA debe ser pública y accesible: en efecto, uno de los principales riesgos a los que como sociedad nos enfrentamos es a que la IA que se utilice en la justicia penal sea configurada, controlada y ejecutada por unos pocos, a mayor abundamiento, del sector privado, y que pudieran obedecer a los intereses de determinados lobbies. Por ello, debiera, de un lado, regularse la IA y, de otro, que sea pública y cualquier ciudadano pueda acceder a ella.[48] Resulta, además, fundamental para el ejercicio del derecho de defensa conocer el algoritmo al efecto de poder impugnar el resultado de la diligencia y/o prueba practicada u obtenida mediante inteligencia artificial.
c. Evaluación y revisión por personas o entidades independientes: la inteligencia artificial se encuentra en constante evolución, por ello deviene imprescindible que su aplicación a la justicia penal sea periódicamente evaluada. Así pues, las herramientas que empleen esta tecnología deberán ser revisadas, principalmente por un grupo de expertos independientes, a ser posible nombrada por una entidad pública, sin perjuicio de que también puedan constituirse grupos o entidades privadas –mejor si éstas son sin ánimo de lucro– a las que se permita una auditoría de la IA para informar sobre las deficiencias y mejoras necesarias de cara a un funcionamiento tanto eficaz como respetuoso con los derechos y garantías del proceso penal.
d. Hacia un relevante papel del Ministerio Fiscal como garante del buen funcionamiento de la IA en el proceso: conviene recordar que el Ministerio Fiscal es, de acuerdo con sus principios de actuación, una parte imparcial en el proceso que debe estar tanto por la condena del culpable como por la absolución del inocente. Es más, la Constitución española, en el artículo 124, le confiere un papel protagonista en la defensa de los

razonamiento y he de expresar mi rotundo rechazo a estos algoritmos predictivos si se interpretan como algo más que un instrumento puramente auxiliar, nunca vinculante –ni siquiera condicionante– al servicio del juez, en quien ha de residenciarse, siempre y en todo caso, la capacidad para afectar la libertad personal de cualquier ciudadano. Sustituir la decisión jurisdiccional por una resolución mecanizada que rinde culto a una supuesta precisión matemática, quebrantaría de modo irreparable las garantías del investigado, de forma especial su derecho de defensa."

48. Armenta Deu, María Teresa, *op. cit.*, p. 319.

derechos de los ciudadanos. De ahí que deba erigirse en el revisor o guardián del correcto uso de la Inteligencia Artificial en el proceso judicial, vigilando su correcto funcionamiento y denunciando las infracciones y vulneraciones de derechos que pueda provocar un uso negligente de esta tecnología.[49]

e. Formación e información para los jueces y operadores jurídicos: la aplicación generalizada de la inteligencia artificial en el proceso sin ni siquiera conocer el fundamento y funcionamiento de esta tecnología puede implicar consecuencias indeseables. Con ello, no se pretende, ni mucho menos, la adquisición de un nivel experto, desde una perspectiva informático-científica, del conocimiento y manejo de esta tecnología. Un jurista no es un informático –ojalá los planes de derecho avancen en ese sentido– pero sí debe, al menos, entender cómo opera la inteligencia artificial. Ya existen actualmente programas muy interesantes sobre *Legal Tech* en los que se puede adquirir una información básica e incluso algo avanzada sobre la aplicación de IA a la justicia penal y, si bien muchos de los destinatarios son abogados ejercientes, también la Escuela Judicial cuenta con un Máster Universitario Oficial para el Ejercicio de la Función Jurisdiccional, impartido por la UNED, en la que los jueces que ingresan en la carrera judicial reciben formación en Inteligencia Artificial, *Blockchain* así como otras tecnologías disruptivas.[50]

f. Mayor pedagogía e información hacia la sociedad: si, tal y como demuestran los índices de referencia, todavía es una asignatura pendiente acercar la Administración de Justicia a la ciudadanía, con mayor razón se les deberá explicar con claridad el papel que tendrá la Inteligencia Artificial en la justicia penal.[51] No se trata, para nada, de una cuestión baladí, pues partimos de una tecnología cuyo desconocimiento generalizado provo-

49. Nieva Fenoll, Jordi, *op. cit.*, p. 150.

50. El plan de estudios, además, incluye seminarios concretos sobre las referidas materias. Disponible en: http://portal.uned.es/portal/page?_pageid=93,70656198&_dad=portal&_schema=PORTAL&idTitulacion=262301 [fecha de consulta: 22/03/2024].

51. El índice de la Unión Europea EU Justice del año 2021. Disponible en: https://ec.europa.eu/info/sites/default/files/eu_justice_scoreboard_2021.pdf [fecha de consulta: 22/03/2024].

ca una lógica desconfianza en una ciudadanía que, a la postre, es la destinataria de las decisiones que, mediante su uso, van a adoptarse en la justicia penal. Ya existen interesantes estudios, destacando el de Morales Moreno, que muestran como la aceptación de esta tecnología y su aplicación a la justicia penal es baja: hoy en día la ciudadanía desconfía de las decisiones judiciales que reposan en predicciones algorítmicas.[52] Por consiguiente, más allá estar incursos en una transformación digital generalizada de la sociedad, deberemos, en particular, ser muy escrupulosos, claros y pedagógicos a la hora de explicar tanto los beneficios como, en definitiva, el funcionamiento, de la aplicación de esta tecnología en la administración de justicia.

Posibles usos futuros

Comenzando por el final, la aplicación de la IA en la justicia penal debe servir al fin de ayudar o asistir al juez en la toma decisiones. Si bien dicha tecnología cuenta con el potencial para, en muchos casos, resultar más fiable que la inteligencia humana, el empleo de algoritmos para la imposición de penas privativas de libertad, al estilo del manido ejemplo de "*COMPAS*" en los EE. UU.,[53] no sería constitucional porque, salvo la conformidad, a nadie se le puede condenar a una pena privativa de libertad sin haber sido sometido a un juicio oral con todas las garantías ante la inmediación de un tribunal imparcial e independiente.[54]

Del mismo modo, también debemos ser precavidos en la aplicación de la IA para la limitación de la libertad en sede cautelar, principalmente a la hora de imponer la prisión provisional. Como sugiere Neira Pena, el juicio de imputación reforzado que debe acontecer ante la imposición de esta medida, no ha de verse contaminado por los sesgos que pudiera adolecer la herramienta, sino que no debe olvidarse que el proceso decisional habrá de desarrollarse de modo individualizado

52. Morales Moreno, África María, "Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales", en *Revista Ius et Scientia*, 2001.

53. Borges Blazquez, Raquel, *op. cit.*, p. 63.

54. Gimeno Sendra, Vicente, *La simplificación de la justicia penal y civil,* BOE, 2020, p. 28.

y en atención a las concretas circunstancias del caso, lo que implica que en último término, con mayor o menor influencia de la máquina, la decisión recaiga en el juez.[55] Pero, sin negar la referida conclusión, sí que es cierto, por otro lado, que intentar objetivar y "algoritmizar" determinados parámetros para la valoración del riesgo de fuga, en un momento como el actual en el que desgraciadamente se abusa de tan restrictiva medida cautelar –la más grave– puede operar a favor del imputado. Como se ha demostrado con el ya analizado RisCanvi, que desde su implementación se conceden más permisos que antes, establecer un sistema de puntuación o *scoring* –p. ej. 3 puntos por arraigo, 5 si justifica que tiene trabajo– puede resultar de gran ayuda para el juez[56] que, sin perjuicio de la debida atención individualizada y a las concretas circunstancias del caso, podrá tomar una decisión que descanse en una base algo más objetiva que lo que sucede en la actualidad.

Asimismo, y dejando a un lado aquello relativo a la limitación de la libertad, todas las propuestas de resolución que impliquen una objetivación o puedan realizarse objetivando parámetros sí podrán aplicar Inteligencia Artificial. Ello ya fue advertido por Gimeno Sendra quien entendió, con buen tino, que la IA podría ayudar o asistir al órgano judicial formulado propuestas de resolución sobre distintos temas.[57] Si bien algu-

55. Neira Pena, Ana María, "Inteligencia artificial y tutela cautelar. Especial referencia a la prisión provisional", en *Revista Brasileira de Direito Processual Penal*, Vol. 7, Nº 3, La Coruña, 2021, p. 1927.

56. Muy gráfico al respecto se muestra Velasco Nuñez, la aplicación de la IA para la valoración del riesgo de fuga "… es la madre del cordero de la prisión provisional. Saber si la persona se va a escapar en el periodo que falta para el juicio o no. Es el sistema 'Scoring'. Con las siguientes variables: 3 puntos si está arraigado, 5 si tiene trabajo […] Cuantos más puntos, menos riesgo de que se vaya a fugar", palabras recogidas en una entrevista publicada en Confilegal el 21/02/2022. Disponible en: https://confilegal.com/20220221-jueces-robot-y-comunicaciones-en-blockchain-una-realidad-cada-vez-mas-cercana-en-espana/ [fecha de consulta: 22/03/2024].

57. Gimeno Sendra, Vicente, *op. cit.*, pp. 28-29. El listado al que se refería, sin ánimo de exhaustividad, Gimeno Sendra, es el siguiente: cómputo del cumplimiento efectivo de las penas (arts. 988. III LECrim y 76 C.P.); Los presupuestos procesales, tales como la jurisdicción (arts.9 y 23 LOPJ) y los conflictos transfronterizos; la competencia objetiva y aforamientos (art. 303 LECrim); competencia sobre conexión de delitos (arts. 17.1.II y 300), competencia territorial (arts.19 y s.s.); la prescripción (arts. 131 y s.s. C.P., 666.3ª LECrim); el perdón del ofendido (arts. 130.1.5º C.P., 106.II LECrim), el indulto (arts. 130.1.4º C.P., 666.4ª LECrim y L. de 18 de junio de 1870) y el suplicatorio (arts. 666.5ª, arts. 11-14 del Reglamento del Congreso de los Diputados y art. 22 del

nos de ellos pueden ser acogidos, entre los que destacaríamos, las medidas cautelares civiles, la fianza del acusador popular y el decomiso; otras propuestas las descartaríamos por su escasa aplicación práctica –p. ej. propuesta de resolución sobre indulto– o incluso por su elevada complejidad y porque pueda comprometer derechos fundamentales tales como la libertad –p. ej. la ejecución de una euroorden de detención y entrega–.

Por consiguiente, debemos huir de planteamientos ambiciosos y partir de un enfoque más prudente en la aplicación de la IA en la justicia penal. Así pues, todo aquello ínsito a la tramitación del procedimiento o incluso al objeto civil –que se acumula al penal ex arts. 100 y 108 LECrim– podría ser un primer estadio en el que aplicar la IA para, posteriormente, y ante una eventual evaluación positiva de su utilización, ampliarlo a otras resoluciones de la justicia penal y, en último término, la asistencia en la valoración de la prueba y en la motivación al órgano judicial. En efecto, la valoración de la prueba ya sea documental, personal o pericial, así como la motivación de la sentencia, podrán ser en un futuro practicadas con ayuda de la inteligencia artificial, pero existirán determi-

Reglamento del Senado), la cosa juzgada (art. 666.2ª).; Las medidas cautelares civiles (arts. 764, 589 y s.s.), la fianza del acusador popular (art. 280) y el decomiso (arts. 127-127 octies C.P.); Las resoluciones provisionales de prohibición de residencia (art. 544 bis LECrim en relación con los arts. 57 y 48 C.P.) y órdenes de protección (art. 544 ter), la privación provisional, en la instrucción, del permiso de conducción (art. 529 *bis* LECrim), la suspensión del funcionario prevista en la legislación administrativa, la suspensión provisional de la función o cargo público del procesado en situación de prisión provisional y sospechoso de pertenecer a una organización terrorista o rebelde (art. 384 *bis* LECrim), la clausura temporal de una empresa y suspensión temporal de las actividades de una sociedad (art. 529.3 C.P.) y el secuestro de publicaciones y la prohibición de difundir las noticias delictivas (arts. 816 y 823 bis) LECrim, 189.8, 270.3 y 510.6 C.P.).; El archivo del atestado por inexistencia de autor conocido (art. 284.2 LECrim).; El sobreseimiento por razones de oportunidad ante los delitos-bagatela (art. 963.1.1ª LECrim).; Los supuestos de mediación penal y/o sentencia de conformidad negociada por razones de oportunidad que, culminadas en la fase intermedia, tengan una tramitación escrita (arts. 784.3 y 787.1); El proceso por aceptación de Decreto (arts. 803 bis.a-803 bis.j); Otras resoluciones: las pruebas del ADN para la determinación del imputado que se efectúan ya mediante algoritmos (arts. 5 L.O. 10/2007 y 3.a del RD 1977/2008) o conjurar el riesgo de reiteración delictiva (art. 129 bis CP), la declaración de rebeldía y de contumacia (arts. 512-514 y 786 LECrim); los plazos de la instrucción (art. 324); la ejecución de una euroorden de detención y entrega (arts. 47 y s.s.) y de una orden europea (arts. 16 y ss L. 23/2014), la petición vinculante de sobreseimiento (arts. 642-645 y 782); resoluciones recurribles (arts. 216 y s.s., 766, 790, 803, 846 bis a), 846 ter, 847 y 976); el depósito del acusador particular (art. 875) y los plazos para la interposición de los recursos (arts. 212, 766.3, 803.1.1ª, 856 y 976.1), etc.

nados parámetros que indefectiblemente requerirán de la presencia humana.[58] Debieran ser, a nuestro juicio, y por la relevancia constitucional de los mismos, los últimos espacios en los que aplicar la IA, una vez haya sido aplicado con éxito en otros actos o resoluciones procesales con una afección menor o menos intensa en los derechos del investigado. Pero, insistimos, ante una justicia penal que todavía se encuentra en plena transformación digital, no podemos pretender un uso generalizado y, lo que es peor descontrolado –esto es, literalmente, carente de control– de la inteligencia artificial. Sólo con unos mimbres sólidos –expuestos en el epígrafe anterior– podrá desarrollarse una aplicación de la IA eficaz, pero a la vez respetuosa con los derechos fundamentales en el proceso penal.

Conclusión

Los instrumentos de policía y justicia predictiva que se utilizan actualmente en la justicia penal no permiten inferir una aplicación intensa y generalizada de esta tecnología en nuestro país. Se impone, por consiguiente, un uso prudente y consonante con lo acontecido en los países y la propia Unión Europea, cuyo Reglamento sobre Inteligencia Artificial todavía está por llegar.[59] Y más allá de que el Gobierno se muestre entusiasta en la transformación digital y destine una gran cantidad de fondos en su plan Justicia 2030, lo cierto es que falta concreción acerca del propósito en la aplicación de la inteligencia artificial en la justicia, más allá de un planteamiento genérico de justicia inteligente orientada al dato.[60] Si no queda claro qué se pretende, es decir, si no se definen con claridad los objetivos en este punto clave de la transformación digital, corremos el riesgo de que, de un lado, se dificulte la innovación y el emprendimiento y, de otro, podamos encontrar tensiones ante iniciativas o soluciones privadas que puedan fracasar en su intento por ser empleadas por las autoridades públicas habida

58. Nieva Fenoll, Jordi, *op. cit.*, p. 87.

59. De Hoyos Sancho, Montserrat, "El uso jurisdiccional de los sistemas de inteligencia artificial y la necesidad de su armonización en el contexto de la Unión Europea", en *Revista General de Derecho Procesal*, Nº 55, Madrid, 2021.

60. El plan de trabajo sobre eficiencia digital. Disponible en: https://www.justicia2030.es/eficiencia-digital [fecha de consulta: 22/03/2024]

cuenta la falta de regulación existente y la afección tanto a la normativa en protección de datos como a los derechos fundamentales.[61]

Del mismo modo, el Poder Judicial parece haber tomado conciencia de la importancia y del impacto que puede tener la Inteligencia Artificial en la administración de justicia. Si bien los sistemas de jurimetría llevan aplicándose desde hace unos cinco años en nuestro país, no ha sido hasta la fecha, en el informe al Anteproyecto –hoy Proyecto– de Ley de Eficiencia Digital, el momento en el que han advertido de la relevancia y riesgo de aplicar Inteligencia Artificial en la justicia penal, siendo necesaria una regulación completa y garantista y erigiéndose en un actor que debe jugar un papel determinante.[62] Más allá de que ese papel determinante le corresponda, a nuestro juicio, al poder legislativo, se creó un grupo de trabajo denominado "Tecnología, Inteligencia Artificial y Administración de Justicia", formado por prestigiosos jueces e ingenieros informáticos que están elaborando un elenco de herramientas de inteligencia artificial para su aplicación en la Administración de Justicia. No obstante lo anterior, y considerando positiva la implicación del CGPJ en el estudio de la IA, desgraciadamente, como suele acontecer

61. Al respecto es importante ver lo que aconteció con los sistemas de reconocimiento facial para la prevención del delito en una conocida cadena de supermercados que fueron duramente cuestionados en el Auto Nº 72/2021 de la Audiencia Provincial de Barcelona, Sección 9ª, Rec Nº 840/2021. En la fundamentación jurídica, la juez alerta que "No todo vale en materia de derechos fundamentales. Estas tecnologías pueden ser realmente intrusivas y requieren de un debate ético y jurídico sosegado, toda vez que pueden tener efectos muy adversos en los valores fundamentales y la integridad humana". Y ello porque –y con esto concluye la fundamentación jurídica– con el reconocimiento facial no se están protegiendo intereses públicos, sino privados de la persona jurídica y "se estarían conculcando las garantías adecuadas en orden a la protección de los derechos y libertades de los interesados, no ya sólo de los que han sido penados y cuya prohibición de acceso les incumbe, sino del resto de personas que acceden al citado supermercado"

62. El informe al Anteproyecto, párrafo 168. Disponible en: https://www.poderjudicial.es/cgpj/es/Poder-Judicial/Consejo-General-del-Poder-Judicial/Actividad-del-CGPJ/Informes/Informe-al-anteproyecto-de-Ley-de-Eficiencia-Digital-del-Servicio-Publico-de-Justicia--por-la-que-se-transpone-al-ordenamiento-juridico-espanol-la-Directiva--UE--2019-1151-del-Parlamento-Europeo-y-del-Consejo--de-20-de-junio-de-2019--por-la-que-se-modifica-la-Directiva--UE--2017-1132-en-lo-que-respecta-a-la-utilizacion-de-herramientas-y-procesos-digitales-en-el-ambito-del-Derecho-de-sociedades [fecha de consulta: 22/03/2024]. Seguidamente, se constata ese enfoque prudente cuando afirman que "Esperar al resultado del procedimiento legislativo de la Unión sobre la propuesta de Reglamento sobre inteligencia artificial es una opción recomendable, antes de abordar la regulación de las denominadas actuaciones asistidas en nuestro ordenamiento jurídico".

ante reformas de tanto calado, una futura reforma que avance hacia un uso más amplio de esta tecnología seguramente no estará exento de resistencias y reticencias por parte de la judicatura.

Así las cosas, es necesario suscribir un enfoque prudente, huyendo de los extremos, sin desdeñar los avances y las oportunidades que presenta esta tecnología, pero, al mismo tiempo, asegurando el respeto a los derechos fundamentales y garantías procesales.[63] No se puede plantear, ni mucho menos, un *totum revolutum*, sino que la aplicación de la inteligencia artificial a la justicia penal debe realizarse de forma paulatina y sosegada, estableciendo revisiones científicas antes de la implementación y después de la misma.[64]Por ello, su implementación no sólo debe ser llevada a cabo por juristas, en un lado, y por informáticos, en otro, sino que, habida cuenta las herramientas deben ser evaluadas y revisadas científicamente, debieran desempeñar un importante papel los criminólogos.

Y por último una reflexión final: si avanzamos hacia una aplicación más intensa de IA en la justicia penal, será necesario superar la frustración que genera el hecho de que la realidad o el resultado alcanzado no se corresponda con la probabilidad de que éste aconteciera. En efecto, conviene recordar que hablamos de justicia predictiva y predecir es sinónimo de pronosticar o adivinar, pero no de infalibilidad. Por ello, reivindicando la belleza de lo improbable, debiéramos naturalizar el extraño y poco frecuente acontecimiento de un resultado distinto al pronosticado por la IA. Que no haya acertado no quiere decir que se haya equivocado –o incluso que esté mal configurada– sino que se ha producido un resultado distinto al pronosticado que no debe conducirnos, inexorablemente, a una conclusión precipitada sobre el mal funcionamiento de esta tecnología. No es aventurado afirmar que la IA será una realidad más pronto que tarde, y aunque de forma asistencial, tendrá cada vez mayor presencia en la justicia penal. No demos la espalda, pues, a una tecnología que, por desconocida, no deja de resultar fascinante y preparémonos hoy para la justicia del mañana.

63. Simón Castellano, Pere, *Justicia cautelar e inteligencia artificial*, Barcelona, Editorial Bosch, 2021, p. 98. En total consonancia con el planteamiento de Simón Castellano, que mantiene una posición "ambivalente" situada "*en el centro de los extremos y que trata de aprovechar las ventajas del avance técnico sin dejar de alertar de los extremos y aristas que esta despliega, fijando ciertas líneas rojas*".

64. Miró Llinares, Fernando,"Predictive policing: utopia or dystopia? On attitudes towards the use of *Big Data* algorithms for law enforcement", en *Revista IDP*, 2020, pp. 1-8.

Workshop: Inteligencia artificial y las reglas del Derecho

Inteligencia Artificial y Estado de Derecho: oportunidades y desafíos

Emmanouil Billis*

Introducción

En nuestro mundo globalizado, los mayores riesgos para la paz, el orden social y la seguridad pública han pasado a ser también globales. En respuesta a los nuevos tipos de amenazas serias y sofisticadas (el terrorismo internacional y el ecológico, la financiación de actividades terroristas, el ciberterrorismo, los nuevos tipos de crimen organizado transnacional y el lavado de dinero), está surgiendo una nueva arquitectura de control de la delincuencia. Se caracteriza por una transformación universal de las nociones jurídicas tradicionales y por la difuminación de los límites entre seguridad y derecho penal, así como entre los conceptos de prevención y represión. La sustitución de los principios de preservación de la libertad personal y la intimidad por formas reforzadas de coerción estatal y vigilancia masiva generalizada es una parte integral de este fenómeno.[1] Gracias a los avances tecnológicos actuales, los mecanismos de represión modernos ya no

* Dr. jur. (Friburgo/Alemania). LL.M. (Bonn/Alemania y Atenas/Grecia). Jefe de Grupo de Investigación en el Instituto Max Planck para el Estudio del Crimen, la Seguridad y el Derecho en Friburgo/Alemania. Abogado ante el Tribunal Supremo Griego. Profesor invitado en la Universidad de Oxford. Investigador invitado en la Queen Mary University de Londres. Profesor invitado en la UiT The Arctic University de Noruega; Miembro de la AIDP.

1. Sin duda, la vigilancia masiva ya está teniendo un amplio impacto en la prevención y represión de la delincuencia: Las medidas de vigilancia excepcionales se han convertido en una supervisión estatal permanente. Y la normalización de las restricciones preventivas, pero todavía masivamente invasivas, de los derechos a la libertad, el movimiento y la intimidad con objetivos de seguridad pública ha transformado sustancialmente los objetivos tradicionales, los métodos operativos y los límites y principios de protección de los sistemas de control de la delincuencia y de justicia penal. Véase Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, "The Typology of Proportionality", en Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter (eds.), *Proportionality in Crime Control and Criminal Justice*, Oxford, Hart Publishing, 2021, pp. 3, 5-11.

apuntan únicamente a los sospechosos o delincuentes tradicionales, sino también a "nuevas" categorías de riesgos para la seguridad: los presuntos sospechosos y los delincuentes potenciales. Estos mecanismos presentan una amplia gama de características intrusivas y coercitivas que pueden afectar a los derechos de sectores mucho más amplios de la población en comparación con el aparato de justicia penal convencional.

Una parte importante de esta evolución se compone por la creciente dependencia de las herramientas de Inteligencia Artificial (IA)[2] para la detección y eliminación automatizadas de amenazas a la seguridad y posibles delitos. Los defensores de estas herramientas aducen la capacidad de la IA para mejorar la aplicación de la ley y la justicia penal de muchas maneras: desde volver más eficaz y eficiente el ejercicio de la autoridad pública y reducir el sesgo de las decisiones jurídicas, hasta limitar el carácter arbitrario de las injerencias estatales en los derechos fundamentales. Sin embargo, el uso de esta tecnología también conlleva muchos problemas vinculados al Estado de Derecho y los derechos humanos. La IA tiene el potencial de ampliar radicalmente los poderes de vigilancia y coerción de los Estados, así como de acelerar drásticamente el proceso de transformación antes mencionado. A medida que introducimos la automatización en nuestras concepciones legales, éstas también se transforman. Uno de los ejemplos más característicos al respecto es el paradigma de la justicia y la vigilancia policial predictiva.

Por lo tanto, los avances tecnológicos en el campo de la IA no sólo presentan grandes oportunidades (en la práctica) para el control de la delincuencia y para la justicia penal, sino también riesgos considerables para la coexistencia pacífica entre seres sociales. La importancia de las consideraciones relativas al Estado de Derecho y a los derechos humanos a la hora de diseñar y aplicar herramientas de IA con fines de seguridad, control de la delincuencia y justicia penal –en particu-

2. Sobre las distintas definiciones de IA, véase, por ejemplo, el art. 3 y anexo I de la "Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial" (Ley de Inteligencia Artificial), 21/04/2021, COM(2021) 206 final; Comisión de la UE, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, "Directrices éticas para una IA fiable", 08/04/2019, p. 36; Comisión de la UE, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, "A definition of AI: Main capabilities and scientific disciplines", 08/04/2019, p. 3 y 7.

lar de nociones clave como la dignidad humana, la proporcionalidad, la igualdad y la justicia procesal– se vuelve así evidente. Por ejemplo, sigue siendo muy dudoso que los algoritmos modernos puedan percibir y emplear con éxito conceptos jurídicos fundamentales, como la equidad y la proporcionalidad, para resultar en una justicia humana. Otros temas de debate importantes en este contexto son los asuntos especialmente delicados, conectados a cuestiones sociojurídicas y éticojurídicas más amplias de la justicia, la legitimidad y la democracia, como las cuestiones de la privacidad, la protección de datos, la seguridad, la fiabilidad, la transparencia y la objetividad, la parcialidad y la discriminación, así como la explicabilidad y la responsabilidad de la IA.[3] Tales consideraciones son pertinentes para una serie de aplicaciones modernas de la IA que sirven a diversos fines: desde la vigilancia policial predictiva, la prevención y la detección de delitos hasta la justicia predictiva, la evaluación del riesgo y de la reincidencia y la determinación del castigo penal.

La investigación y la política jurídicas se enfrentan a la necesidad inmediata de encontrar formas de abordar en la práctica los desafíos relevantes. El intercambio continuo de conocimientos entre los juristas y los informáticos es una condición previa fundamental para desarrollar estrategias eficaces con el fin de alcanzar un entendimiento mutuo sobre la aplicación real de las nuevas tecnologías en el mundo jurídico e idear un plan sobre cómo traducir con éxito las nociones jurídicas tradicionales y los principios de protección al lenguaje de la programación. Una de las primeras prioridades en este contexto debe ser facilitar el diseño y el funcionamiento de algoritmos y máquinas de acuerdo con los objetivos primordiales de proteger y garantizar el respeto de los valores

3. Comisión de la UE, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, "Directrices éticas para una IA fiable", 08/04/2019, p. 2: "Una IA confiable tiene tres componentes, que deben alcanzarse durante todo el ciclo de vida del sistema: (1) debe ser legal, cumpliendo con todas las leyes y reglamentos aplicables (2) debe ser ética, adhiriendo a los principios y valores éticos y (3) debe ser robusta, tanto desde una perspectiva técnica como social, ya que, incluso con buenas intenciones, los sistemas de IA pueden causar daños involuntarios. (...) [Debe garantizarse] que el desarrollo, despliegue y uso de los sistemas de IA cumplan los siete requisitos clave para una IA digna de confianza: (1) agencia y supervisión humanas, (2) solidez y seguridad técnicas, (3) privacidad y gobernanza de datos, (4) transparencia, (5) diversidad, no discriminación y equidad, (6) bienestar medioambiental y social y (7) responsabilidad".

humanos y sociales más básicos. En particular, en los ordenamientos democráticos liberales que emplean la tecnología de la IA, la atención debe centrarse en garantizar los requisitos constitutivos del Estado de Derecho: el principio de legalidad (incluida la aplicación coherente e imparcial de normas e instituciones previsibles, claras y transparentes); los principios de igualdad y proporcionalidad; el uso no arbitrario del poder; el respeto de los derechos fundamentales y las garantías procesales; la separación de los poderes del Estado y el control de su ejercicio por órganos judiciales independientes e imparciales.[4]

El enfoque centrado en el ser humano

En cuanto a la relación entre la IA y el Estado de Derecho específicamente, una tarea importante y un desafío es programar proactivamente las herramientas algorítmicas de forma de excluir la arbitrariedad en los procesos de toma de decisiones que se basen en tales herramientas. En segundo lugar, también necesitamos optimizar el funcionamiento y los procesos de aprendizaje de la IA con el propósito general de ayudar y complementar a la justicia tradicional en la producción de resultados más precisos, objetivos y justos. Especialmente en los ámbitos de la seguridad y el control de la delincuencia, los sistemas del Estado de Derecho tienen el deber negativo de señalar y prohibir el uso de aplicaciones que supongan una amenaza directa para la dignidad humana, así como de programar los algoritmos de forma que se garantice su respeto.[5] Al

4. "Rule of Law Checklist", adoptada por la Comisión de Venecia del Consejo de Europa en su 106ª sesión plenaria (Venecia, 11-12 de marzo de 2016), párrafos 9 y ss., 15-18, 31 y ss.; Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo y al Consejo, "Further strengthening the Rule of Law within the Union", 03/04/2019, COM (2019) 163 final, p. 1.

5. Para la consideración de la dignidad humana como base real de los derechos fundamentales, que sitúa al individuo en el centro de la actuación de los Estados, véase la Carta de los Derechos Fundamentales de la UE (explicaciones). Sobre el reconocimiento de la dignidad como algo inviolable y absoluto para proteger a los seres humanos de ser menospreciados o tratados arbitrariamente como meros objetos por el Estado, véanse las conclusiones del Tribunal Constitucional alemán en los casos BVerfGE 27, 1 (6) [1969] y BVerfGE 30, 1 (25-26) [1970]. Sobre el "derecho madre" a la dignidad humana; Barak, Aharon, *Human Dignity: The Constitutional Value and the Constitutional Right*, Cambridge, Cambridge University Press, 2015, pp. 156-167. Sobre la dignidad humana como "la libertad de configurar la propia vida", Dupré, Catherine, "Article

mismo tiempo, también tienen el deber positivo derivado del objetivo primordial de cualquier evolución tecnológica que es respetar, proteger y promover los valores humanos fundamentales: desarrollar máquinas y sistemas electrónicos que puedan ayudar a salvaguardar al individuo frente a las acciones coercitivas de los órganos estatales que sean arbitrarias, crueles, inclementes o desproporcionadas.

Así, de los valores fundamentales interconectados de la dignidad humana, la autonomía, la libertad y el Estado de Derecho se desprende, en primer lugar, la necesidad de un enfoque centrado en el ser humano para todo lo vinculado con el funcionamiento y el uso de la IA.[6] Los elementos básicos de este enfoque centrado en el ser humano o que prioriza al ser humano son:[7]

- El establecimiento de fuertes controles y garantías adicionales para usos específicos de la IA, especialmente las aplicaciones utilizadas en la toma de decisiones y la aplicación de leyes que tengan un alto potencial para dañar a las personas.
- La identificación de usos de la IA ética y socialmente cuestionables, y su clasificación como inaceptables y prohibidos (como en el caso de los sistemas de IA que otorgan puntaje social a personas físicas).
- La regulación exhaustiva del desarrollo y del uso de aplicaciones consideradas de "alto riesgo" y más bien peligrosas, intrusivas y problemáticas, *entre otras*, por cuestiones

1 - Human Dignity", en Peers, Steve, et al. (eds.), *The EU Charter of Fundamental Rights. A Commentary*, Oxford, Hart Publishing, 2ª ed., 2021, pp. 3, 6.

6. Sobre el enfoque centrado en el ser humano, véase Comisión de la UE, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, "Directrices éticas para una IA fiable", 08/04/2019, p. 37: "Una IA con un enfoque centrado en la persona se esfuerza por asegurar que los valores humanos ocupen un lugar central en el desarrollo, despliegue, utilización y supervisión de los sistemas de IA, garantizando el respeto de los derechos fundamentales, todos ellos constituyen una referencia unitaria a un fundamento común arraigado en el respeto de la dignidad humana, en el que el ser humano disfruta de una condición moral única e inalienable".

7. Arts. 5 y ss, "Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)", 21/04/2021, COM (2021) 206 final; y "Resolución del Parlamento Europeo sobre la inteligencia artificial (IA) en el derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales", de 06/10/2021 (2020/2016(INI)), pp. 16, 17, 32.

discriminatorias, parcialidades o falta de transparencia (ejemplos característicos son el reconocimiento facial en lugares públicos y los sistemas predictivos de vigilancia).

- El pleno respeto de los derechos a la intimidad, la libertad de circulación, la presunción de inocencia y los derechos de defensa (derecho al silencio, libertad de expresión e información, igualdad ante la ley, igualdad de armas, derecho a un recurso efectivo y a un juicio justo). Lo mismo ocurre con la explicabilidad algorítmica, la transparencia, la trazabilidad y la supervisión efectiva.[8]

Desde este punto de vista, no debemos considerar el enfoque centrado en el ser humano como un obstáculo para el progreso, sino más bien como una garantía para la coexistencia armónica en nuestra sociedad tecnológica global. A veces simplemente es necesario instaurar mecanismos de supervisión e intervención, requisitos estrictos de funcionamiento y utilizar restricciones amplias e incluso prohibiciones totales sobre determinadas aplicaciones de la IA con el fin de garantizar eficazmente la dignidad, la autonomía y la libertad. Sin embargo, algunos de los requisitos de protección, normas del Estado de Derecho y garantías de los derechos humanos propuestos en los diversos textos jurídicos contemporáneos e instrumentos del *soft-law* pueden no ser tan fáciles de aplicar y cumplir en la práctica, al menos en el estado actual de la evolución tecnológica y jurídica. Todavía hay cuestiones sin resolver que deben abordarse con urgencia en los debates políticos y en la investigación interdisciplinar, entre las cuales están los problemas de cómo desarrollar conjuntos de datos de entrenamiento libres de errores o del riesgo de sesgos; cómo crear sistemas de aprendizaje automático totalmente explicables y transparentes (independientemente de su complejidad interna o de las barreras externas a la transparencia planteadas por las empresas tecnológicas y los intereses de la propiedad intelectual); cómo producir decisiones y

8. Particularmente importante es la rendición de cuentas y el elemento de la intervención humana con respecto a todas las aplicaciones policiales de la IA, específicamente en el sentido de que "todas las decisiones con efectos legales deben ser tomadas siempre por un ser humano al que puedan pedirse cuentas de las decisiones adoptadas" ("Resolución del Parlamento Europeo sobre la inteligencia artificial (IA) en el derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales", de 06/10/2021 (2020/2016(INI)), p. 16.

resultados automatizados objetivos y no discriminatorios; y cómo diseñar modelos funcionales de rendición de cuentas y responsabilidad. Por último, al considerar cuestiones relevantes también desde una perspectiva filosófica o ética, no deberíamos pasar por alto los diversos escenarios distópicos que hemos experimentado recientemente, especialmente durante la crisis global de Covid-19, con los que debemos lidiar como sociedades organizadas. Esto incluye, por ejemplo, los sistemas de crédito social puestos en marcha en algunos países,[9] mecanismos de predicción muy avanzados ya en desarrollo o simplemente el problema general de cómo aceptar que los robots "imiten" a los humanos encargados de hacer cumplir la ley en un mundo por lo demás centrado en el ser humano.

El principio de proporcionalidad

En sintonía con esta evolución, una de las primeras nociones que vienen a la mente, que es crucial para el fortalecimiento del enfoque centrado en el ser humano y, en general, para el mantenimiento del Estado de Derecho, es el concepto de proporcionalidad, un concepto de importancia diacrónica, especialmente para el derecho penal y de seguridad en términos de su función protectora contra las injerencias arbitrarias sobre los derechos individuales y las libertades adquiridas.[10] Hoy en día, la importancia de la proporcionalidad ha aumentado

9. Vogler, Richard, "*Big Data* and Criminal Justice. Proportionality, Efficiency and Risk in a Global Context", en Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, (eds.), *Proportionality in Crime Control and Criminal Justice*, Oxford, Hart Publishing, 2021, pp. 165, 174 y ss.

10. Comisión de la UE, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, "Directrices éticas para una IA fiable", 08/04/2019, pp. 12-13; "Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre inteligencia artificial (Ley de Inteligencia Artificial)", 21/04/2021, COM (2021) 206 final, pp. 7, 11, 21 y ss; "Resolución del Parlamento Europeo sobre la inteligencia artificial (IA) en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales", de 06/10/2021 (2020/2016(INI)). La proporcionalidad configura y limita el ejercicio de todos los poderes del Estado. Es una técnica de razonamiento jurídico, un método moderado de control de la autoridad pública y, al mismo tiempo, un factor racional para mejorar la aceptación social de medidas coercitivas e intrusivas pragmáticamente necesarias en los derechos - esto significa que la proporcionalidad también puede contribuir a garantizar sistemas jurídicos más

aún más con respecto al uso de las nuevas tecnologías de vigilancia masiva e inteligencia artificial con fines de seguridad y lucha contra la delincuencia.

La tarea de considerar la proporcionalidad en el contexto de la IA tiene dos niveles. En primer lugar, con respecto al uso proporcionado de la IA: los legisladores deben ser capaces de definir criterios claros y no arbitrarios y los órganos administrativos y judiciales deben ser capaces de ejercer autocontrol y control para garantizar el empleo legítimo, adecuado, menos intrusivo y apropiado de la IA en las diversas constelaciones. En segundo lugar, en lo referido a los propios resultados de la IA y el aprendizaje automático, que deben ser proporcionales: El principal desafío consiste en desarrollar los algoritmos y "entrenar" a las máquinas para que puedan tomar decisiones y emprender acciones de conformidad con los elementos básicos de la proporcionalidad.[11]

Una vez más, desde un punto de vista práctico, en el estado actual del progreso tecnológico, es más fácil decirlo que hacerlo, velar por el desarrollo de máquinas que sean realmente capaces de "decidir", "actuar" y "reaccionar" de forma templada y proporcionada eligiendo los medios más adecuados, menos intrusivos y apropiados para alcanzar objetivos legítimos de seguridad y control de la delincuencia. Y, también en este asunto, hay objeciones desde puntos de vista filosóficos y éticos. La más importante es la cuestión de si los cálculos y las decisiones sobre la proporcionalidad o la excesividad

humanos y funcionales a largo plazo, véase en este sentido Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, "The Typology of Proportionality" en Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter (eds.), *Proportionality in Crime Control and Criminal Justice*, Oxford, Hart Publishing, 2021, pp. 3, 11 y ss.

11. En el marco de las evaluaciones de proporcionalidad, los intereses en conflicto se ponderan en los tres niveles diferentes del poder público, afectando a medidas y decisiones de los tres poderes: el legislativo, el ejecutivo y el judicial. En el núcleo de estas evaluaciones se encuentra el examen de los siguientes elementos generalmente reconocidos: la existencia de un objetivo legítimo que justifique la adopción de una determinada medida; la idoneidad de la medida para alcanzar este objetivo concreto; la necesidad de esta medida particular a la vista del objetivo (requisito de la medida menos intrusiva); y la idoneidad de la medida en términos de sus efectos (limitadores de derechos) sopesados con los beneficios del objetivo perseguido (proporcionalidad en el sentido estricto). Para más detalles, Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, "The Typology of Proportionality" en Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter (eds.), *Proportionality in Crime Control and Criminal Justice, op. cit.*, p. 24.

de una determinada acción pueden realmente ser el resultado de la aplicación de fórmulas mecánicas. ¿O deben esas tareas permanecer como "juicio de valor" tomado por un actor humano responsable "de buena fe y de manera razonable"?[12]

Un buen ejemplo en el ámbito del derecho humanitario internacional y el derecho penal internacional es el uso de sistemas de armas autónomas (AWS por sus siglas en inglés) en situaciones de conflicto armado. En efecto, los procesadores y algoritmos de los sistemas de armas autónomos pueden ser capaces de evaluar la información recopilada para llevar a cabo la selección de objetivos e identificar los daños incidentales esperados de un posible ataque militar sin verse influidos por el estrés y otras emociones humanas, por lo que están libres del riesgo de error humano. Por lo tanto, los AWS pueden ser (o, en algún momento, serán) capaces de tomar decisiones más objetivas. Sin embargo, los procesos matemáticos y las evaluaciones de proporcionalidad realizados por AWS son más bien lineales. Se basan sobre todo en parámetros técnicos y prefijados para la recolección de información mediante el uso de sensores y en la rígida interpretación y aplicación de normas jurídicas previamente definidas por procesadores y algoritmos. Pero los AWS carecen todavía de inteligencia contextual y de la capacidad humana de comprender y adaptarse a realidades en constante cambio, utilizando inferencias aprendidas que permiten a los individuos percibir sus propias acciones también desde la perspectiva de los demás. Por lo tanto, los AWS actuales que operan sobre la base de algoritmos predeterminados no pueden garantizar realmente que la toma de decisiones militares se traduzca siempre en la elección de los medios y objetivos adecuados, menos intrusivos y más apropiados, como exige el principio de proporcionalidad.[13]

12. Bothe, Michael y Gillard, Emanuela-Chiara, en Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter (eds.), *Proportionality in Crime Control and Criminal Justice*, *op. cit.*, p. 295.

13. Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, "Künstliche Intelligenz und der Grundsatz der Verhältnismäßigkeit", en Engelhart, Marc; Kudlich, Hans; Vogel, Benjamin (eds.), *Digitalisierung, Globalisierung und Risikoprävention - Festschrift für Ulrich Sieber zum 70. Geburtstag. Geburtstag, Teilband II*, Berlín, Duncker & Humblot, 2021, pp. 715 y ss.; Billis, Emmanouil y Knust, Nandor, "Proportionality (Principle of)", en Caiero, Pedro et al. (eds.), *Elgar Encyclopedia of Crime and Criminal Justice*, Cheltenham, Edward Elgar Publishing, 2023.

Consideraciones similares pueden aplicarse en general a los avances tecnológicos modernos que permiten la recolección automatizada, la selección y el tratamiento "en tiempo real" de grandes volúmenes de datos personales y a los sistemas electrónicos altamente sofisticados que ayudan en los procesos de toma de decisiones pertinentes. Las consideraciones éticas y los desafíos pragmáticos para conseguir resultados proporcionales de la IA en los ámbitos de la seguridad, el control de la delincuencia y la justicia penal no cambian, por supuesto, el hecho de que la proporcionalidad sigue siendo una importante herramienta de limitación y control contra cualquier ejercicio arbitrariamente coercitivo e intrusivo de los poderes del Estado. No obstante, a la luz del dominio tecnológico actual en la vigilancia masiva y otros ámbitos de intrusión en la intimidad, la proporcionalidad es "solo una" entre las muchas preocupaciones relacionadas con el Estado de Derecho que la teoría, la práctica y la política tienen el deber de abordar. Especialmente la interferencia con derechos de privacidad bien establecidos –derechos que no son absolutos, pero cuya restricción sólo está permitida por razones excepcionales y debe estar sujeta a límites estrictos– mediante un uso desenfrenado y no supervisado de la IA puede socavar las limitaciones de protección, las salvaguardas democráticas y las garantías del Estado de Derecho existentes.

En estos momentos, es mucho lo que está en juego para el Estado de Derecho y la protección efectiva de los derechos de autodeterminación y privacidad. Los datos personales constituyen el elemento fundamental de las aplicaciones basadas en la IA utilizadas con fines de seguridad y control de la delincuencia. Estas aplicaciones no están exentas de los riesgos de manipulación, discriminación y falta de transparencia ni de los problemas de supervisión y rendición de cuentas insuficientes. La principal preocupación con respecto al uso de algoritmos para producir o adquirir información sensible y "nuevos conocimientos" recae sobre el conocido problema de la "caja negra" y la posibilidad de que redunde en correlaciones imprevisibles e inferencias inexplicables. Además, los algoritmos actuales se entrenan y funcionan principalmente sobre la base de observaciones y correlaciones estadísticas.[14] Como resultado,

14. Sgaier, Sema K.; Huang, Vincent; Charles, Grace, "The Case for Causal AI", en *Stanford Social Innovation Review*, 50: "Por muy sofisticados que sean, los algoritmos

siempre existe el peligro de que la información y los conocimientos obtenidos por los sistemas algorítmicos actuales estén distorsionados o sean discriminatorios y poco fiables.[15]

Los principios de igualdad y justicia procesal

En particular, el problema de los resultados posiblemente discriminatorios constituye una preocupación habitual en el contexto de la toma automatizada de decisiones y las predicciones de la IA. Este problema depende principalmente de factores como la calidad de los datos, los prejuicios preexistentes introducidos en el sistema algorítmico por sus diseñadores y formadores y/o la incapacidad del sistema para desarrollar un razonamiento verdaderamente causal.[16] La prohibición general de discriminación se deriva del principio de igualdad. Tanto la igualdad como la no discriminación son parte de los fundamentos del Estado de Derecho. Además, los derechos de igualdad y la prohibición de la discriminación están relacionados con la exigencia primordial de respetar y proteger la dignidad humana. En términos de la aplicación de la ley y justicia penal, concretamente, el riesgo de que los sistemas algorítmicos generen resultados discriminatorios suele debatirse en el contexto de la identificación de personas, la vigilancia predictiva y las aplicaciones de evaluación del riesgo de reincidencia.[17] Aunque en

predictivos y sus usuarios pueden caer en la trampa de equiparar correlación con causalidad", Stanford, Stanford Center on Philanthropy and Civil Society, 2020, p. 18.

15. Billis, Emmanouil; Knust, Nandor; Rui, Jon Petter, "Künstliche Intelligenz und der Grundsatz der Verhältnismäßigkeit", en Engelhart, Marc; Kudlich, Hans; Vogel, Benjamin (eds.), *Digitalisierung, Globalisierung und Risikoprävention - Festschrift für Ulrich Sieber zum 70. Geburtstag, Teilband II*, *op. cit.*, pp. 706-707.

16. "Resolución del Parlamento Europeo sobre la inteligencia artificial (IA) en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales", de 06/10/2021 (2020/2016(INI)), pp. 8, 22, 23, 24. Como se ha señalado, las decisiones algorítmicas "pueden terminar amplificando y reproduciendo sesgos existentes", véase Ryberg, Jesper y Roberts, Julian, "Sentencing and Artificial Intelligence - Setting the Stage", en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence*, Oxford, Oxford University Press, 2022, pp. 1, 8.

17. "Resolución del Parlamento Europeo…", *op. cit.*, pp. 9, 24, 27; Angwin, J.; Larson, J.; Mattu, S.; y Kirchner, L., "Machine Bias", en *ProPublica*. Disponible en: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [fecha de consulta: 23/05/2016]; Davies, Benjamin y Douglas, Thomas, "Learning to

tales constelaciones la IA puede promover efectivamente la igualdad de trato y la imparcialidad, su uso también puede menoscabar estos valores sociales y jurídicos fundamentales.[18]

En este contexto, un aspecto adicional del Estado de Derecho que hay que considerar es la relación entre la igualdad y la justicia procesal. Los elementos esenciales de la justicia procesal son la imparcialidad, la congruencia, la precisión y la veracidad, la representatividad, la adecuación ética, la transparencia, la posibilidad de revisar y recurrir una decisión y el trato respetuoso de los afectados por el conflicto y el procedimiento.[19] Aunque, en ocasiones, estos elementos pueden entrar en conflicto entre sí (más aún en el ámbito de la IA y la justicia penal),[20] en los sistemas democráticos liberales la justicia y la legitimidad social requieren, como mínimo, la aplicación de normas formales, garantías procesales adecuadas y derechos procesales justos que proporcionen oportunidades sustanciales y equitativas de inclusión, participación y recurso procesales.[21]

En sintonía con esto, el uso de la IA debe respetar siempre los principios de igualdad ante la ley e igualdad de armas, así como los derechos a un recurso efectivo y a un juicio justo. Especialmente en la

---

Discriminate - The Perfect Proxy Problem in Artificially Intelligent Sentencing", en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence* Oxford, Oxford University Press, 2022, p. 97; Lippert-Rasmussen, Kasper, "Algorithm-Based Sentencing and Discrimination", en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence, op. cit.*, p. 74.

18. Bagaric, Mirko y Hunter, Dan, "Enhancing the Integrity of the Sentencing Process through the Use of Artificial Intelligence" en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence, op. cit.*, pp. 122 y ss., 131 y ss.

19. Rehbinder, Manfred, *Rechtsoziologie*, 6ª ed., Munich, Beck, 2007, p. 118.

20. Véase, por ejemplo, sobre la posibilidad de una relación negativa entre la transparencia de un algoritmo y su precisión en la imposición de penas (algoritmos complicados y opacos que producen predicciones más precisas de la criminalidad futura *versus* algoritmos que son más transparentes, pero producen predicciones menos precisas), Ryberg, Jesper y Petersen, Thomas, "Sentencing and the Conflict between Algorithmic Accuracy and Transparency", en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence, op. cit.*, p. 57.

21. Billis, Emmanouil y Knust, Nandor, "Alternative Types of Procedure and the Formal Limits of National Criminal Justice: Aspects of Social Legitimacy" en Sieber, Ulrich; Mitsilegas, Valsamis; Mylonopoulos, Christos; Billis, Emmanouil y Knust, Nandor, (eds.), *Alternative Systems of Crime Control. National, Transnational, and International Dimensions*, Berlín, Duncker & Humblot, 2018, pp. 39, 57.

aplicación de la ley y la justicia penal, la tecnología de la IA podría servir como medio novedoso y alternativo para procesar/producir automáticamente información y pruebas con el fin de mejorar la eficacia, la objetividad, la imparcialidad y la economía procesal. Sin embargo, hay una preocupación pragmática importante que mencionar a este respecto: Normalmente existirá una asimetría de poder entre quienes emplean las tecnologías de IA y quienes están sometidos a ellas, algo que, si no prestamos la debida atención, podría incluso llevar a que la IA produzca mayor desigualdad, división social o exclusión.[22] Por lo tanto, una prioridad importante de los sistemas del Estado de Derecho en cuanto a asuntos de justicia y procedimiento penal, como las pruebas y las sentencias, sería limitar el uso de la IA con la aplicación de salvaguardas y derechos suficientes para impugnar la toma automatizada de decisiones por motivos de igualdad, discriminación y justicia procesal.

Aquí, el ejemplo de las sentencias penales es característico. Un objetivo ampliamente reconocido para el empleo de herramientas de IA en la imposición de penas es garantizar no sólo decisiones rápidas, sino también más objetivas e imparciales, libres de las "restricciones" de las emociones humanas, los prejuicios y los intereses personales.[23]

22. "Resolución del Parlamento Europeo…", (2020/2016(INI)), *op. cit.*, pp. 2 y 10, también sobre el impacto del uso de herramientas de IA sobre los derechos de defensa de los sospechosos, la dificultad de obtener información significativa sobre su funcionamiento y la consiguiente dificultad para impugnar sus resultados ante los tribunales, en particular por parte de las personas investigadas. El Parlamento Europeo (p. 16) ha señalado además que a) en contextos judiciales y policiales, la decisión que produce efectos jurídicos o similares debe ser tomada siempre por un ser humano, que pueda ser considerado responsable de las decisiones adoptadas; b) las personas sometidas a sistemas basados en IA deben poder recurrir las sentencias; c) una persona tiene derecho a no ser objeto de una decisión que produzca efectos jurídicos que le conciernan o le afecten de manera significativa y que se base únicamente en un tratamiento automatizado de datos; d) la toma de decisiones individuales automatizadas no debe basarse en categorías especiales de datos personales, a menos que existan medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado; e) está prohibida la elaboración de perfiles que dé lugar a la discriminación de personas físicas sobre la base de categorías especiales de datos personales; f) las autoridades que utilicen sistemas de inteligencia artificial deben respetar normas jurídicas muy estrictas y garantizar la intervención humana, especialmente cuando analicen datos procedentes de dichos sistemas, y g) es necesario respetar la discreción soberana de los jueces y la toma de decisiones caso por caso.

23. La teoría y la práctica reciente en diversos ordenamientos jurídicos han demostrado que los nuevos tipos de IA y aprendizaje automático –elogiados por su presunta

Sin embargo, las capacidades de las nuevas tecnologías para reforzar la administración de la justicia penal y del Estado de Derecho no deben eclipsar los posibles riesgos para la igualdad, la objetividad y la imparcialidad causados por la propia IA. Las máquinas pueden utilizarse para apoyar al sentenciador humano en sus tareas. Pero entonces surge la pregunta: ¿hasta qué punto pueden los jueces, consciente o inconscientemente, seguir siendo verdaderamente independientes e imparciales? Si los jueces confían demasiado en la superioridad técnica de las máquinas o se apoyan excesivamente en ellas, resulta inevitable que basen sus decisiones en la información procesada y los conocimientos generados por algoritmos.[24] Sin embargo, como ya se ha visto, los resultados de los sistemas algorítmicos actuales pueden estar distorsionados, ser discriminatorios y poco fiables por consecuencia de sesgos sistemáticos o de deficiencias técnicas en el razonamiento causal. Por lo tanto, es necesario diseñar y aplicar rígidamente salvaguardas procesales adecuadas para garantizar, a todas las partes interesadas, la igualdad de posibilidades de participación, la simetría de la información y los recursos, la apertura y la transparencia en los procesos conjuntos de toma humana y automatizada de decisiones y, en consecuencia, oportunidades adecuadas de revisión y apelación. El objetivo es lograr un efecto de compensación de las deficiencias inherentes a las aplicaciones de la IA en la justicia penal y, al mismo tiempo, un efecto de fortalecimiento de la aceptación social del resultado final.

---

capacidad matemática para reducir la complejidad fáctica y mejorar la calidad (coherencia, precisión, objetividad) y la rapidez en la toma de decisiones, pero también criticados por plantear riesgos potenciales para los principios protectores tradicionales y las limitaciones del Estado de Derecho– ya están transformando el proceso penal convencional. Véanse los diversos análisis en Ryberg, Jesper y Roberts, Julian (eds.), *Sentencing and Artificial Intelligence, op. cit.*, 2022.

24. "Resolución del Parlamento Europeo...", (2020/2016(INI)), *op. cit.*, p. 15: "... [Si] los seres humanos se basan únicamente en datos, perfiles y recomendaciones generados por máquinas, no podrán realizar una evaluación independiente; resalta las consecuencias negativas potencialmente graves, particularmente en el ámbito de las actividades policiales y judiciales, que pueden derivarse de una confianza excesiva en la naturaleza aparentemente objetiva y científica de las herramientas de IA, sin tener en cuenta la posibilidad de que sus resultados sean incorrectos, incompletos, irrelevantes o discriminatorios; hace hincapié en que debe evitarse el exceso de confianza en los resultados ofrecidos por sistemas de IA y destaca la necesidad de que las autoridades adquieran confianza y conocimientos para poner en cuestión recomendaciones algorítmicas".

Conclusión

En las modernas sociedades del riesgo, la delincuencia se ha vuelto técnica, compleja y transnacional, mientras que los ordenamientos jurídicos se han vuelto pesados y están sobrecargados. Esto ha acrecentado la importancia práctica de emplear mecanismos alternativos en la lucha contra las amenazas a la seguridad y la delincuencia, y en la administración de la justicia penal. Como resultado, ha surgido una cantidad de nuevas aplicaciones tecnológicas, en paralelo a las prácticas tradicionales de aplicación de la ley y la justicia, con el objetivo de reforzar la eficacia y la eficiencia de los sistemas de control de la delincuencia y de la justicia penal. Constantemente se desarrollan nuevas técnicas de lucha contra la delincuencia y nuevas herramientas para la administración de la justicia y para mejorar las funciones preventivas y represivas de las fuerzas policiales, así como para simplificar y volver más eficaces las investigaciones y los juicios penales.

La aplicación de políticas y normativas transparentes, coherentes y orientadas según los derechos humanos en el campo requiere un conocimiento suficiente del funcionamiento interno de estas tecnologías altamente sofisticadas desde la teoría y la práctica jurídicas. Al mismo tiempo, los informáticos y desarrolladores de programas informáticos también deben tener un conocimiento firme de los principios y conceptos jurídicos que subyacen al funcionamiento de los sistemas jurídicos contemporáneos. Un requisito previo es la identificación precisa, la evaluación y, en caso necesario, la modelización sistemática de las ideas organizativas, los razonamientos y los factores que determinan las estructuras básicas de los mecanismos e instituciones contemporáneos de prevención, investigación, enjuiciamiento y resolución de conflictos. Además, al explorar las oportunidades, los beneficios prácticos y los retos de los modernos métodos e instrumentos de vigilancia e investigación de la Inteligencia Artificial, en constante evolución, todas las partes implicadas también deben tener en cuenta los considerables riesgos para la coexistencia pacífica de los seres humanos y el Estado de derecho asociados a este tipo de avances tecnológicos.

Esto significa que la política y la investigación no deben centrarse únicamente en regular y aplicar eficazmente las tecnologías modernas para combatir la delincuencia. Por el contrario, deben hacer especial

hincapié en garantizar las nociones de estado de derecho, legitimidad social y derechos humanos. Al menos en principio, nuestros sistemas de Estado de Derecho siguen sujetos a las mismas obligaciones: prevenir y reprimir las conductas ilícitas, aspirar a la resolución veraz y justa de los conflictos, preservar la economía procesal y la paz social, así como respetar los valores y principios básicos de la dignidad humana, la proporcionalidad y la clemencia. La investigación comparativa e interdisciplinaria de los fundamentos, centrada en estos elementos es, por tanto, esencial para la resolución eficaz de los problemas y las reformas políticas, en el ámbito del derecho y la tecnología más que nunca.

Diacrónicamente, los planificadores de estrategias, los políticos y las empresas tecnológicas, así como los organismos encargados de hacer cumplir la ley y los servicios de inteligencia, parecen haber dado más importancia a las nuevas puertas que se abren y a las oportunidades que ofrecen tecnologías como la IA en los ámbitos de la aplicación de la ley y la justicia. Por otro lado, la teoría y la investigación jurídicas suelen hacer hincapié en los posibles peligros que un uso descontrolado y no supervisado de la IA puede causar para la intimidad, los derechos de protección de datos y los requisitos del estado de derecho en materia de fiabilidad, transparencia, objetividad, no discriminación y rendición de cuentas. El hecho de que, comparado con los avances tecnológicos anteriores, el empleo de la IA en el control de la delincuencia y la justicia penal pueda dar lugar a amenazas más amplias, inmediatas y de múltiples niveles para los derechos y libertades establecidos, vuelve necesaria, según algunos expertos, una prohibición amplia de aplicaciones específicas, que se identifican como altamente intrusivas y particularmente peligrosas para las personas y la sociedad. A pesar de esta preocupación, en el presente capítulo de se ha argumentado que, en consonancia con la aparente inevitabilidad de la expansión de la IA, la atención principal debe centrarse en construir los algoritmos y programar las máquinas de acuerdo con las exigencias normativas de cuatro principios fundamentales: la dignidad humana, la proporcionalidad, la igualdad y la justicia procesal.

Renegociar el contrato social. La Inteligencia Artificial en la vigilancia predictiva y su impacto en la legitimación del control social a través de los poderes coercitivos del Estado

Nandor Knust*

Introducción

La función general del derecho penal es ser el instrumento *ultima ratio* del Estado para garantizar el orden y la paz sociales. La razón de ser del poder público, es decir, la salvaguarda de la libertad y la dignidad humana de todos los individuos, puede discernirse en las nociones de contrato social y Estado de Derecho, que definen y rigen nuestra existencia social en comunidad y sociedad.

El contrato social es un acuerdo para formar una entidad que, por definición, es algo más que una mera agregación de intereses y voluntades individuales. Mediante la renuncia colectiva a los derechos y libertades de que goza el individuo en el "estado de naturaleza" y la transferencia de estos derechos al ente colectivo, se forma una nueva "persona" (soberano/estado). Esta versión del contrato social incluye la idea de deberes recíprocos: El soberano/estado se compromete con el bien de los individuos y, a su vez, cada individuo se compromete con el bien común. Esta construcción social está respaldada y controlada por el concepto de Estado de Derecho. Las autoridades públicas de los Estados de derecho tienen el mandato de salvaguardar la libertad y la dignidad humana de todos los individuos.

En las sociedades modernas, el control social y la paz se mantienen gracias a los sistemas policiales y de justicia penal, autoridades

* Profesor asociado de la UIT. Universidad Ártica de Noruega.

que toman decisiones con consecuencias serias para las libertades y la conducta de las personas. Recientemente, estas decisiones se ven cada vez más influenciadas por algoritmos especiales destinados a predecir conductas delictivas y por las correspondientes acciones policiales preventivas. El creciente uso de *software* predictivo por parte de la policía crea el riesgo de un uso generalizado y descontrolado de datos y metadatos.[1] La integración de algoritmos incomprensibles y cada vez más autónomos en los sistemas de toma de decisión de las fuerzas del orden y la justicia penal amenaza seriamente las ideas antes mencionadas del contrato social y el Estado de Derecho. Estos nuevos algoritmos son capaces de decidir por sí mismos qué y cómo aprenden y seleccionan, lo que los vuelve no solo opacos para las personas sin los conocimientos tecnológicos necesarios, sino que, debido a su inmensa complejidad en el proceso de toma de decisiones, pueden incluso constituir una caja negra para sus desarrolladores.[2] Esta evolución supone un cambio profundo en los procesos de toma de decisión de la policía y de la justicia y, por lo tanto, modifica también la relación entre la policía y la sociedad.[3]

Las características de este nuevo tipo de algoritmos plantean varias cuestiones sobre el papel de la inteligencia artificial en el control social, pero también en el cambio de las expectativas normativas y la previsibilidad de la acción estatal. Los procesos automatizados de toma de decisiones de este tipo tienen implicaciones enormes para la organización estructural de las instituciones para la prevención del delito, control de la delincuencia y justicia penal. Con este telón de fondo, este capítulo analiza si el cambiante panorama del control de la delincuencia por medio de la vigilancia predictiva requiere una "renegociación" del contrato social entre el Estado, como detentor del monopolio del poder, y sus ciudadanos.

1. Mozur, Paul; Xiao Muyi y Liu, John, "An Invisible Cage: How China Is Policing the Future", en *The New York Times*, 26/06/2022. Disponible en: https://www.nytimes.com/2022/06/25/technology/china-surveillance-police.html

2. Esposito, Elena, "Transparency Versus Explanation: The Role of Ambiguity in Legal AI", en *Journal of Cross-disciplanary Reseach in Computational Law*. Disponible en: https://journalcrcl.org/crcl/article/view/10/8.

3. Egbert, Simon; Esposito, Elena; Heimstätt, Maximilian, "Vorhersagen und Entscheiden: Predictive Policing in Polizeiorganisationen", en *Soziale Systeme*, 2021, pp. 26, 189 y 191.

El Derecho y el contrato social

El Derecho es una parte esencial en la estructura de los Estados nacionales democráticos y en el diseño del orden social en la sociedad. La sociedad se basa en la comunicación como unidad de enunciado, información y comprensión, y está conformada por comunicaciones contingentes. El Derecho y otros sistemas sociales se diferencian funcionalmente unos de otros por el uso de su propio código binario para diferenciarse dentro de sus fronteras funcionales y delimitar otros sistemas sociales.[4] Así, los sistemas producen y dan forma a sus propios elementos por sí mismos según sus propios códigos binarios de diferenciación y, como tales, operan (comunicándose) de forma autónoma.[5] El sistema jurídico funciona por sí mismo diferenciando si algo es legal o ilegal. Este código de diferenciación y el hecho de que el Estado sea el legítimo detentor del monopolio del poder para garantizar el orden y la paz social permiten a las fuerzas policiales utilizar su poder para intervenir en las esferas protegidas de los ciudadanos. Cuando el ordenamiento jurídico define una actividad observada como un acto potencialmente "ilegal", el Estado puede utilizar su poder para hacer frente a la posible fechoría con el fin de restablecer el orden y la paz sociales. Esta intervención del Estado mediante el uso de la fuerza está regulada por un complejo sistema de controles y equilibrios, en el que se sopesan las libertades individuales y la obligación del Estado de garantizar la paz y la estabilidad. Pero, aunque el sistema jurídico funcione por sí solo, puede verse influido por otros sistemas y su comunicación a través de acoplamientos estructurales.[6] Por lo tanto, la sociedad y sus sistemas sociales están estimulados por varias fuentes diferentes y, en última instancia, se basan en su propia contingencia.[7]

La ley y su legitimidad se basan en procesos democráticos y en el sistema de libertad individual y derechos humanos. Como ya se ha

4. Nobles, Richard y Schiff, David, "Using Systems Theory to Study Legal Pluralism: What Could Be Gained", en *Law & Soc'y Rev*, 2012, pp. 265-270.

5. Andersen, Simon Calmar, "How to Improve the Outcome of State Welfare Services. Governance in A Systems-Theoretical Perspective", en *Public Administration*, 2005.

6. Ídem.

7. King, Michael y Thronhill, Chris, *Luhmann on Law and Politics: Critical Appraisals and Applications*, Oxford, Hart Publishing, 2006, p. 8.

mencionado, una pluralidad de sistemas interactivos se comunica entre sí en el seno de la sociedad, configurándola mediante sus interacciones. Por un lado, las normas jurídicas impregnan la sociedad y se convierten en parte integrante de las esferas política y social, con una profunda influencia en los ámbitos público y privado de la sociedad. Por otro lado, el Derecho y su función deben ajustarse a nuevas irritaciones, impulsos y retos como los cambios institucionales debidos a la creciente especialización y a una complejidad cada vez mayor en las sociedades modernas impulsadas por la tecnología. Estos ajustes constantes se plasman en un sistema generalizado de comunicación, que facilita las interacciones del sistema jurídico con otros sistemas sociales.[8] Las sociedades modernas se basan en la comunicación en tiempo real y se diferencian por funciones sociales que se expresan comunicativamente. Por lo tanto, la sociedad continua y constantemente conformada e impulsada por el desarrollo de sistemas comunicativos y procesos específicos y basados en el conocimiento, la toma de decisiones y la discursividad.[9] Esto es muy importante para entender el uso de la información y la comunicación basado en el *software* de vigilancia predictiva. Estos programas informáticos producen constantemente información que las fuerzas del orden utilizan e integran en su propia comunicación. Y el sistema jurídico utiliza esta información para decidir sobre posibles actividades policiales.

Esta evolución no es exclusiva del sistema jurídico. Más generalmente, puede observarse que el discurso dentro de las sociedades se está volviendo muy tecnológico y que los dispositivos tecnológicos son partes (semi)autónomas de esta comunicación. El auge de sistemas de libre acceso como ChatGPT está produciendo cambios drásticos en las interacciones sociales, las comunicaciones y los intercambios al construir y proporcionar contenido propio dentro de la comunicación. Estas tendencias también afectan a la comunicación dentro del sistema jurídico y, por tanto, podrían provocar una desestabilización de sus

8. Sand, Johanne Inger, "Changing Forms of Governance and the Role of Law - Society and its Law, Arena Working Paper", *University of Oxford*. Disponible en: https://www.sv.uio.no/arena/english/research/publications/arena-working-papers/1994-2000/2000/00_14.html [fecha de consulta 05/04/2024]

9. Ídem.

procesos internos.[10] La externalización de la toma de decisiones a especialistas o instituciones especializadas se ha convertido en una parte esencial del sistema jurídico.[11] La información que debe procesarse en la toma de decisiones jurídicas es cada vez más compleja y adopta la forma de discursos fragmentados y plurales, que dependen en parte de puntos de vista externos y de la participación de otros actores, lo que conduce a una mayor reflexividad.[12] Esta complejidad, basada en el cambio constante y la integración de las nuevas tecnologías en el discurso, conduce a un aumento de la incertidumbre, que está influyendo en la racionalidad de las decisiones jurídicas. Como resultado, el sistema jurídico se ve desestabilizado y cuestionado por las formas de comunicación tecnológica dentro y fuera del sistema, lo que provoca graves consecuencias para las funciones del sistema jurídico, así como para su interacción con los individuos y otros sistemas sociales.[13] Tomando eso en cuenta, ¿qué significa para el contrato social, la vigilancia predictiva y el uso de la Inteligencia Artificial?

La ley es un actor importante en el contrato social entre el Estado y los individuos porque sirve para generar, satisfacer y estabilizar las expectativas normativas de la sociedad y para resolver conflictos basados en la ley y su aplicación. El uso de la aplicación de la ley es la herramienta *ultima ratio* del Estado como propietario del monopolio del poder para hacer cumplir las reglas normativas y garantizar el orden social. Sin embargo, en la actual sociedad del riesgo empujada por la tecnología, las nuevas innovaciones e impulsos tecnológicos no siempre se incorporan firmemente al ordenamiento jurídico ni se traducen adecuadamente en normas o reglamentos jurídicos específicos, ni el diseño actual de la infraestructura jurídica puede responder eficazmente a los desafíos sociales actuales.[14] La extensión de nuestra realidad e interacción social hacia un entorno digital dificulta que el sistema jurídico genere y estabilice expectativas, procese la enorme

10. Ídem.

11. Knust, Nandor, *Socio-legal foundation for public-private partnerships in the system of anti-money-laundering and counterterrorism financing*, Propuesta para PartFin Project.

12. Giddens, Anthony, *The Consequences of Modernity*, Stanford, Stanford University Press, 1990.

13. Sand, Johanne Inger, *op. cit.*

14. Ídem.

cantidad de información adicional que este entorno digital sigue produciendo y resuelva nuevas formas de conflictos. Además, no existe un intercambio automático entre las esferas social y jurídica, sino que siempre se requiere la traducción a un código que el otro sistema pueda entender.[15] Sin embargo, aunque una de las funciones centrales del Derecho en el contrato social es reducir la complejidad de las distintas esferas sociales para incrementar la previsibilidad y predictibilidad, no cualquier impulso o irritación puede traducirse siempre al código jurídico y conducir a la creación y satisfacción de expectativas normativas.[16] Pero si el derecho ha de cumplir su función de generar y satisfacer expectativas normativas en la sociedad, sus operaciones deben estar limitadas por las funciones y racionalidades específicas de la ley.[17]

La complejidad es una característica clave de la toma de decisiones en las sociedades modernas en términos de decisiones diversas que deben tomarse simultáneamente, creando así una pluralidad de irritaciones incontrolables multiplicadas por las comunicaciones e interacciones entre los diversos responsables y actores de la sociedad.[18] La acumulación de un gran número de interacciones constituye la complejidad. La contingencia es otro efecto secundario inevitable de la acumulación de decisiones selectivas.[19] Las decisiones de gran complejidad y contingencia pueden conducir en última instancia a una mayor opacidad e inestabilidad.[20]

El auge de las nuevas tecnologías y del procesamiento, la diferenciación y la comunicación digitales afecta fuertemente al sistema jurídico y a la aplicación de la ley, de modo que las demandas de coherencia y simplicidad del sistema jurídico y su razonamiento son cada vez más difíciles de satisfacer.[21] Esto también afecta a la toma de deci-

15. El concepto de traducción/diccionario entre juristas e informáticos centrado en el uso de las nuevas tecnologías para el control de la delincuencia es el principio rector para el grupo de investigación sobre control de la delincuencia y seguridad de la UIT. Disponible en: https://uit.no/research/crimecontrol [fecha de consulta 05/04/2024].

16. Sand, Johanne Inger, *op. cit.*

17. Ídem.

18. Luhmann, Niklas, *Das Recht der Gesellschaft,* Frankfurt, Suhrkamp, 1995, p. 288 y ss.

19. Luhmann, Niklas, *Ausdifferenzierung des Rechts,* Frankfurt, Suhrkamp 1981, p. 200 y ss.

20. Sand, Johanne Inger, *op. cit.*

21. Ídem.

siones jurídicas dentro del propio sistema jurídico, así como al uso de información procedente de fuera del sistema jurídico que se introduce en él y también en el sistema policial por medio de la obtención de pruebas.[22] En la actualidad, el sistema jurídico se enfrenta a tal sobrecarga (tan cuantitativa como cualitativa) de información que está a punto de alcanzar sus límites funcionales.[23] Una forma de responder a estos límites funcionales es trasladar la resolución de conflictos a otros sistemas mediante su externalización, como puede observarse en el contexto de los sistemas de sanciones administrativas, o en el uso de nuevas formas de toma de decisiones, como resolución alternativa o informal de conflictos.[24] Las nuevas tecnologías, como los programas informáticos de toma de decisiones o de vigilancia predictiva, también se utilizan cada vez más para ampliar los límites funcionales, manejar la enorme masa de información existente y hacerla utilizable para el sistema jurídico y policial.

Este desarrollo plantea un desafío extremo a la racionalidad del sistema de aplicación de la ley, dado su fuerte enfoque en la previsibilidad, la transparencia, la estabilidad y la independencia y autonomía del sistema jurídico respecto a otros sistemas.[25] En términos más generales, la reflexividad interna de la ley requiere aceptar elementos discursivos más diversos, complejos y cambiantes en la ley y, por lo tanto, hacer frente a una creciente inestabilidad, opacidad e incertidumbre como parte de sus patrones internos de razonamiento y dinámica, lo que conduce a un aumento de los riesgos sociales en la aplicación de la ley.[26] Esto, sin embargo, plantea varios interrogantes, especialmente

22. South China Morning Post, "Chinese Scientists Develop AI 'Prosecutor' That Can Press Its Own Charges", 26/12/2021. Disponible en: https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own [fecha de consulta 05/04/2024].

23. Sieber, Ulrich, "The New Architecture of Security Law - Crime Control in the Global Risk Society", *Alternative Systems of Crime Control*, Berlin, Duncker, Humblot, 2018, pp. 3-38.

24. Para más información, véase el Grupo de Investigación Otto Hahn sobre Sistemas Alternativos e Informales de Control de la Delincuencia y Justicia Penal. Disponible en: https://csl.mpg.de/research-groups/crime-control-and-criminal-justice [fecha de consulta 05/04/2024].

25. Knust, Nandor, *"Not criminal responsible in Norway - a brief overview of Section 20 and July 22"*, Canadá, en *Canadian Justice Report*, 2023.

26. Sand, Johanne Inger, *op. cit.*

en lo que respecta al derecho penal y a la aplicación de la ley, ya que se tiene en alta estima salvaguardar el régimen de los derechos humanos y las normas mínimas de protección de los ciudadanos frente a un Estado excesivamente dominante.

Orden social, proporcionalidad y contrato social

Un deber negativo fundamental de los Estados constitucionales liberales y de las comunidades internacionales es limitar las medidas coercitivas oficiales que restringen la libertad individual y significan una amenaza directa para la dignidad humana. Esto se logra consagrando la definición de los derechos individuales fundamentales y humanos en términos constitucionales y legales. Al mismo tiempo, los Estados y las instituciones democráticas modernas tienen el deber positivo de garantizar activamente el ejercicio irrestricto de estos derechos. Esto se traduce a menudo en la necesidad de introducir medidas de seguridad y coercitivas para prevenir las amenazas contra la libertad y la dignidad humana y proteger al público. Además de los instrumentos preventivos de control de la delincuencia, los sistemas de justicia penal deben estar diseñados para hacer frente, de manera eficaz y justa, a cualquier injerencia grave en las libertades, los derechos humanos individuales y los derechos jurídicos colectivos de otras personas.[27]

Las garantías fundamentales de los derechos humanos, los mecanismos eficaces de seguridad y prevención y el funcionamiento de los sistemas de justicia penal contribuyen indudablemente a asegurar la paz social en el mundo globalizado de hoy. Pero el deber positivo de proteger activamente la libertad y la dignidad humana suele entrar en conflicto con el deber negativo de salvaguardar las libertades humanas de la usurpación por parte de los poderes del Estado. Esto requiere un equilibrio constante entre los deberes positivos y negativos del Estado. La concepción y la aplicación del principio de proporcionalidad

27. Billis, Emmanouil; Knust, Nandor y Rui, Jon-Petter, "Künstliche Intelligenz und der Grundsatz der Verhältnismäßigkeit", en *Digitalisierung, Globalisierung und Risikoprävention: Festschrift für Ulrich Sieber zum 70. Geburtstag* Vol.2, Berlin, Duncker & Humblot, 2021, pp. 693-703.

desempeñan un papel vital en la búsqueda de los medios adecuados para alcanzar este equilibrio.[28]

El principio de proporcionalidad y las garantías de justicia y derechos humanos son piedras angulares del Estado constitucional democrático y social. Al mismo tiempo, la cuestión de la eficacia es un elemento esencial de las estrategias actuales en la lucha contra la delincuencia, los mecanismos de resolución de conflictos y los modelos de sanción.[29] Eficacia y proporcionalidad pueden entenderse como conceptos a la vez opuestos e interdependientes.[30] Los avances tecnológicos acelerados de la sociedad global del riesgo actual son un desafío para la prevención y persecución del crimen. Esto lleva a cuestionar los conceptos tradicionales de eficacia/eficiencia y proporcionalidad y, por lo tanto, los modelos de control de la delincuencia deben ajustarse u otros nuevos deben crearse.[31] No obstante, siguen entrando en juego los dos problemas clave del principio de proporcionalidad que son bien conocidos: la falta de una definición precisa y concluyente de sus elementos individuales y su inconmensurabilidad. [32]

Cuando un Estado interfiere en la libertad de una persona, está obligado a actuar dentro de los estrictos límites del sistema del Estado de Derecho. En primer lugar, las intervenciones en las libertades y los derechos individuales de una persona deben estar reguladas por ley.[33] La segunda norma básica es el requisito de proporcionalidad. Las aplicaciones de IA se basan principalmente en el análisis de datos asistido por algoritmos, lo que requiere la recopilación de grandes cantidades de datos en bruto, lo que llamamos *Big Data*. La recopilación de datos personales por parte de las autoridades estatales en el contexto del control de la delincuencia puede constituir una colisión con el Art. 8

28. Ídem.

29. Ídem.

30. Ídem.

31. Ídem.

32. Duff, Anthony, "Proportionality and the Criminal Law: Proportionality of What to What?", en *Proportionality in Crime Control and Criminal Justice*, Oxford, Hart Publishing, 2021.

33. Dorsen, Norman, *Comparative Constitutionalism: Cases and Materials*, 4ta edición, Eagan, West Group, 2003, pp. 43-44.

de la Convenio Europeo de Derechos Humanos (CEDH).[34] "Vida privada" incluye cualquier información relativa a una persona física identificada o identificable. El Tribunal Europeo de Derechos Humanos ha dictaminado que no sólo los datos personales directos entran en el ámbito de protección del Art. 8; basta con que la información sea suficientemente detallada como para crear o establecer la identidad de la persona.[35] Por tanto, toda información que conduzca o pueda conducir a la identificación de una persona se considera dato personal. De ahí que incluso la información pública recolectada y almacenada sistemáticamente en archivos por las autoridades públicas pueda caer dentro de la protección de la "vida privada".[36] En el caso de los sistemas policiales predictivos, esto puede significar que la recogida y almacenamiento de datos, que es un requisito previo para cualquier tipo de sistema policial predictivo, ya constituye una injerencia en la vida privada según el artículo 8 de la CEDH.

El análisis de datos personales o datos personalizados constituye una injerencia sustancialmente diferente en la intimidad comparada con la recolección y almacenamiento de datos personales.[37] La primera diferencia y la más significativa entre la recolección y almacenamiento y el análisis de datos es que –basándose en las correlaciones– el propio análisis "produce" información nueva y previamente desconocida sobre el individuo.[38] Además, el análisis de datos personales conocidos también tiene el potencial de crear y revelar nuevos datos personales.[39] La nueva información sobre una persona puede ser un dato personal sensi-

34. Brinkhoff, Sven, "*Big Data* Data Mining by the Dutch Police: Criteria for a Future Method of Investigation", en *European Journal for Security Research* 2, 2017, pp. 57 y 60.

35. TEDH, P. & S. / Polonia, Nº 57375/08, 30/10/12, ap. 130.

36. TEDH, Shimovolos / Rusia, Nr. 30194/09, 21/06/2011, ap. 55.

37. Buckley, Chris; Mozur, Paul y Ramzy, Austin, "How China Turned a City Into a Prison", en *The New York Times*, 2019. Disponible en: https://www.nytimes.com/interactive/2019/04/04/world/asia/xinjiang-china-surveillance-prison.html [fecha de consulta 05/04/2024]

38. Billis, Emmanouil; Knust, Nandor y Rui, Jon-Petter, *op. cit*.

39. Temme, Merle, "Algorithms and Transparency in View of the New General Data Protection Regulation", en *European Data Protection Law Review* 3, 2017, pp. 473-478; Wachter, Sandra y Mittelstadt, Bernt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Data and AI", en *Columbia Business Law Review* 2019, pp. 494, 559, 577, 615 y 617.

ble, especialmente cuando los datos pueden establecer un vínculo entre una persona y un acto delictivo, sentando así las bases para la apertura de una investigación por parte de las fuerzas de seguridad. Aquí, el problema de la incertidumbre y la falta de previsibilidad desempeñan un papel vital. Los sistemas modernos de IA y sus algoritmos producen nuevos conocimientos y una pluralidad de realidades, especialmente si estos sistemas utilizan *Big Data* para crear contenidos. La cuestión de hasta qué punto puede entenderse la producción de este contenido es vital para el sistema de control de la delincuencia y sus destinatarios. Si el sistema funciona de forma incomprensible[40] –es decir, que no es transparente para el público en general ni explicable para los expertos– se suele calificar como un "dilema de caja negra".[41] El hecho de que los algoritmos operen a nivel estadístico –y no a nivel de causa y efecto– agrava el problema e incluye el riesgo de que la información obtenida por los algoritmos sea sesgada, discriminatoria y poco fiable.[42] Las acciones estatales intrusivas resultantes de la aplicación del poder y la violencia contra un individuo, como las investigaciones penales, las detenciones y las sanciones, podrían, por tanto, resultar en intervenciones e injerencias graves en la esfera privada de los individuos.

Siguiendo la idea del contrato social, el Estado solo puede intervenir en las esferas protegidas de sus ciudadanos en situaciones

40. Ohm, Paul, "Changing the Rules: General Principles for Data Use and Analysis", *Privacy, Big Data and the Public Good: Frameworks for Engagement*, 2014, pp. 100; Burell, Jenna, "How the machine 'thinks': Understanding opacity in *machine learning* algorithms", en *Big Data & Society*, 2016, pp. 1-10. Disponible en: https://doi.org/10.1177/2053951715622512 [fecha de consulta 05/04/2024]; Kuner, Christoph y otros, "The Rise of Cybersecurity and Its Impact on Data Protection", en *International Data Privacy Law*, 2017, pp. 73 -75; Rademacher, Timo, "Predictive Policing im deutschen Polizeirecht", *AöR*, 2017, pp. 366, 376-377; Kuner, Christopher y otros, "Expanding the artificial intelligence-data protection debate", en *International Data Privacy Law 8*, 2018, pp. 289, 290-291; Wischmeyer, Thomas, "Regulierung intelligenter Systeme", *AöR*, 2018, pp. 1, 42-65.

41. Rubinstein, Ira S., "*Big Data*: The End of Privacy or a New Beginning", *International Data Privacy Law 3*, 2013, pp. 74-76; Mittelstadt, Brent Daniel, "The ethics of algorithms: Mapping the debate", en *Big Data & Society 3*, 2016, p.6. Disponible en: https://doi.org/10.1177/2053951716679679 [fecha de consulta 05/04/2024]; Wachter, Sandra y Mittelstadt, Bernt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Data and AI", en *Columbia Business Law Review*, 2019, pp. 494-502.

42. Zhabina, Alena, "How China's AI Is Automating the Legal System", 01/20/2023. Disponible en: https://www.dw.com/en/how-chinas-ai-is-automating-the-legal-system/a-64465988 [fecha de consulta 05/04/2024].

específicas, que deben estar claramente establecidas por ley. Tanto la recolección y almacenamiento de *Big Data* como su análisis son "intrusiones" en la vida privada de un individuo y, por tanto, deben tener un fundamento jurídico estricto. Para cumplir esta obligación derivada del principio de legalidad, no es necesaria una ley parlamentaria. Lo importante aquí es el principio de previsibilidad: garantizar que la intervención estatal sea previsible para el individuo.

El principio de proporcionalidad comprende varios elementos que se expresan con pequeñas variaciones en los distintos ordenamientos jurídicos. Su aplicación también puede diferir en cierta medida, por ejemplo, en función del elemento al que se conceda mayor peso. No obstante, esto parece ser más una cuestión de conceptualización y aplicación práctica que de contenido y realidades. Se pueden identificar cuatro puntos clave para cumplir con los requisitos de una prueba de proporcionalidad en el Estado de Derecho:

- la injerencia debe perseguir un fin legítimo,
- debe ser adecuada y necesaria para este fin,
- la intervención (cumple la condición anterior) debe sopesarse con el derecho humano a proteger,
- la necesidad legítima de la injerencia debe tener más peso que el derecho humano objeto de la injerencia (adecuación o proporcionalidad en un sentido estricto).

Sin embargo, el uso de la IA en el control de la delincuencia –especialmente con programas incomprensibles– choca con los límites del modelo tradicional de proporcionalidad.

Aplicación de la ley, nuevas tecnologías y contrato social

Las fuerzas policiales son organizaciones formales con mecanismos de toma de decisiones propios, que tienen un enorme impacto sobre los individuos y la sociedad en general.[43] La gravedad de sus decisiones y la ejecución de sus acciones están facultadas y legitimadas por el contrato social entre el Estado y sus ciudadanos. Como se ha destacado anterior-

43. Egbert, Simon; Esposito, Elena; Heimstätt, Maximilian, "Vorhersagen und Entscheiden: Predictive Policing in Polizeiorganisationen", en *Soziale Systeme*, 2021, pp. 189-190.

mente, las fuerzas del orden aplican medidas preventivas generadas por estrategias de toma de decisiones que utilizan nuevas tecnologías con algoritmos especiales para predecir conductas delictivas.

Este uso de programas informáticos específicos para la actuación policial predictiva está en consonancia con la observación generalizada de que el desarrollo general del análisis algorítmico de datos mediante técnicas de aprendizaje automático y la disponibilidad de *Big Data* están cambiando el complejo social tanto de las sociedades nacionales como de la comunidad internacional. Algunos de estos nuevos sistemas pueden decidir por sí mismos qué y cómo aprenden y se desarrollan.[44] Estos algoritmos de aprendizaje automático tienen la capacidad de definir o modificar las reglas de toma de decisiones de forma autónoma.[45] Esta interacción entre el uso de meta conjuntos de datos y el aprendizaje profundo crea un complejo proceso de toma de decisiones.[46] Debido a su complejidad y a su autorreferencia cerrada, estos sistemas son poco transparentes e incomprensibles para los humanos. Incluso los informáticos, incluidos los desarrolladores de los algoritmos, pueden no comprender cómo procede la máquina y sobre qué premisas toma las decisiones.[47] Las características de este nuevo tipo de algoritmos plantean interrogantes sobre las implicancias de esta incomprensibilidad para la toma de decisiones en los organismos de vigilancia.

Para una mejor comprensión –y para una mejor discusión de los diferentes modelos de prevención hacia el final de este capítulo– es necesario distinguir dos variantes de predicción algorítmica de la delincuencia.[48] Siguiendo a Egbert, Esposito y Heimstädt, el criterio clave para la distinción es el grado de comprensión que pueda alcanzarse de

44. Burrell, Jenna, "How the Machine 'Thinks': Understanding Opacity in *Machine learning* Algorithms", *Big Data & Society*, 2016, pp. 3, 1, 5 y ss. Disponible en: https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512 [fecha de consulta 08/04/2024].

45. Mittelstadt, Brent Daniel, "La ética de los algoritmos: Mapping the debate", *Big Data & Society*, 2016, pp. 1-3. Disponible en: https://doi.org/10.1177/20539517166796 [fecha de consulta 08/04/2024].

46. Burrell, Jenna, *op cit.*

47. Esposito, Elena, "Transparency Versus Explanation: The Role of Ambiguity in Legal AI", *Journal of Cross-disciplanary Reseach in Computational Law 1*, pp. 1-2. Disponible en: https://journalcrcl.org/crcl/article/view/10/8 [fecha de consulta 08/04/2024]

48. Egbert, Simon; Esposito, Elena y Heimstätt, Maximilian, *op. cit.*, pp. 189, 191 y ss.

las predicciones del programa.[49] Lo decisivo es en qué medida las personas que deciden sobre la difusión y la aplicación de las predicciones están suficientemente informadas sobre las razones del procedimiento, sus criterios y el resultado creado.[50]

- *Predicción comprensible:* Es un tipo de vigilancia predictiva que utiliza programas de pronóstico que generan predicciones inteligibles. La comprensibilidad se basa principalmente en el uso de datos específicos de delitos y en una base teórica relativamente simple. La comprensibilidad de estas predicciones resulta del hecho de que las predicciones se corresponden en gran medida con los conocimientos previos de las fuerzas del orden.[51]
- *Predicción incomprensible:* El programa de predicción incomprensible, por el contrario, se caracteriza por una pluralidad de fuentes, como podrían ser diferentes teorías criminológicas, así como distintos métodos de aprendizaje automático para la predicción de probables actividades delictivas. Otra diferencia clave es que estos sistemas informáticos acceden no solo a datos específicos de delitos, sino también a muchos otros datos no específicos de delitos del conjunto de *Big Data*.[52] Como resultado, este nuevo sistema informático no sólo excede "el conocimiento de la aplicación de la ley y el programa de toma de decisiones", sino también las fuentes de datos tradicionales sobre la aplicación de la ley, al utilizar toda la información comprendida como *Big Data*.[53]

Dentro del concepto de contrato social, las partes/actores estatales implicados deben utilizar el monopolio del poder de forma comprensible y transparente, y sus acciones deben ser eficaces y proporcionadas. El sistema social de las fuerzas policiales consiste a través de la

49. Ídem.

50. Ídem.

51. Ibídem, p. 192.

52. Ibídem, p. 194.

53. Custers, Bart, "New Digital Rights: Imagining Additional Fundamental Rights for the Digital Era", en *Computer Law & Security Review*, 2022. Disponible en: https://doi.org/10.1016/j.clsr.2021.105636 [fecha de consulta 08/04/2024].

comunicación,[54] que conduce a las decisiones[55] y excluye simultáneamente otras posibilidades. La selección y toma de tales decisiones son transparentes y están directamente vinculadas al órgano decisorio. Como resultado, se genera una forma de absorción de la incertidumbre que se reproduce en las operaciones del sistema, lo que produce nuevas decisiones.[56] De este modo, la contingencia y la arbitrariedad potencial de las decisiones se ven ceñidas a premisas para la toma de decisiones o programas de toma de decisiones que limitan la gama de opciones entre las que se puede elegir en un momento dado.[57] Para mantener la actividad dentro de esta gama de opciones, los organismos encargados de hacer cumplir la ley cuentan con normas y procedimientos específicos que determinan qué tipo de actividad es correcta o incorrecta.[58] Dado que las organizaciones funcionan mediante decisiones, los programas de toma de decisiones determinan sus condiciones de corrección. En el caso de las organizaciones, los programas de decisión pueden ser programas *condicionales* o programas de *propósito*.

- *Los programas condicionales* adoptan la forma de "si-entonces", lo que significa que una condición que debe cumplirse para poner en marcha una determinada secuencia de decisiones:[59] Si se denuncia un delito, las fuerzas del orden pueden iniciar una investigación.
- *Los programas de propósito* se centran en un objetivo concreto y, por tanto, legitiman los medios necesarios para alcanzarlo:[60] Para crear un barrio más seguro, es necesario instalar más cámaras de vigilancia.

Obviamente, estas distinciones no siempre son claras, sino a menudo más bien borrosas y, por tanto, artificiales, y, sin embargo, la separación en estos dos tipos diferentes de programas y su abordaje de

54. Luhmann, Niklas, *Die Gesellschaft der Gesellschaft*, Frankfurt, Suhrkamp, 1997, p. 70.

55. Ibídem, p. 830 y ss.; Luhmann, Niklas, *Organisation und Entscheidung*, Bäcker, Dirk, *Opladen, Westdeutscher Verlag*, 2000, p. 63.

56. Luhmann, Niklas, *Organisation und Entscheidung*, Bäcker, Dirk, *op. cit.*

57. Ibídem, p. 222 y ss.

58. Luhmann, Niklas, *Das Recht der Gesellschaft, op. cit.*, p. 208 y ss.

59. Ibídem, p. 263 y ss.

60. Luhmann, Niklas, *Organisation und Entscheidung*, *op. cit.*, p .265 y ss.

la incertidumbre vinculada al futuro es necesaria como herramienta de observación a efectos de análisis.

La programación condicional sigue la estructura de que toda actividad que no esté claramente marcada como permitida por el propio programa está prohibida:[61] Las fuerzas del orden sólo pueden "interferir" en los derechos de libertad individual de los ciudadanos si existen indicios claros de un posible delito y, por tanto, de una conducta punible. Aunque sigue habiendo incertidumbre sobre las actividades futuras de las fuerzas del orden, también se proporciona una cierta estructura previsible debido a las indicaciones claras sobre cuándo y cómo las fuerzas del orden pueden utilizar su poder/violencia contra sus ciudadanos. El Estado está limitado por las normas y garantías mínimas de sus ciudadanos, que sólo pueden violarse en circunstancias previamente establecidas, para garantizar el orden social y la paz para la sociedad en general. Esto explica cierta previsibilidad y predictibilidad de las intervenciones del Estado en forma de actividades policiales.

Por el contrario, los *programas de propósito* siguen la idea de que todo lo que no está estrictamente prohibido está permitido:[62] Las fuerzas del orden disponen de cierta flexibilidad dentro de los límites de su jurisdicción para experimentar, desarrollar y aplicar medios y métodos para reducir la delincuencia en un territorio determinado. En este caso, la incertidumbre sobre el futuro se mantiene, pero, en comparación con *los programas condicionales*, es casi desestructurada. La predictibilidad y previsibilidad de la acción estatal en forma de aplicación de la ley son casi imposibles.

En una combinación de los dos tipos de programas mencionados, las fuerzas policiales son capaces de utilizar las experiencias pasadas manteniendo al mismo tiempo una actitud abierta hacia el futuro. Este enfoque retrospectivo y prospectivo suele entenderse como la naturaleza represiva y preventiva de las actividades policiales. Al combinar ambos programas, las fuerzas policiales disfrutan de una flexibilidad mucho mayor. Sin embargo, esto es a costa de eludir las normas mínimas y el régimen de protección que el contrato social introdujo en el Estado democrático moderno. La separación entre actividades represivas y

61. Ibídem, p. 263.
62. Ibídem, p. 266.

preventivas es formalmente muy estricta, ya que tiene una importancia jurídica considerable si una actividad policial se lleva a cabo por motivos preventivos o represivos.[63] Como ya se ha mencionado, esta división casi puede equipararse a la distinción entre programas *condicionales* y programas de *propósito*, porque la mayoría de las actividades represivas caben bajo los programas *condicionales* y la mayoría de las actividades preventivas pueden resumirse dentro de los programas de *propósito*.

Las fuerzas policiales centradas principalmente en actividades represivas están estrictamente reguladas por programas *condicionales*: El agente policial sólo puede intervenir en determinadas esferas protegidas si una conducta determinada se clasifica como desviada. En estos casos concretos, las fuerzas de seguridad no tienen margen de maniobra, por lo que la innovación en el ejercicio de las actividades represivas será considerada una violación de los derechos de los ciudadanos.[64] Esta prohibición de innovación espontánea para las fuerzas policiales puede considerarse la manifestación de la previsibilidad y transparencia de la actividad estatal por parte del propietario del monopolio del poder.

A diferencia de los organismos encargados de la aplicación de la ley que operan principalmente en el ámbito de las operaciones preventivas, que se centran en tareas mucho más generales guiadas por *programas de propósito* e intentan enfrentar el futuro. La prevención parte de la premisa de que puede producirse un delito, pero no está claro quién, cuándo y dónde se cometerá. Así, las actividades preventivas no se desarrollan en intervenciones precisas y predefinidas y se producen de forma flexible e impredecible.[65]

Si los programas informáticos de predicción algorítmica elaboran predicciones sobre dónde puede producirse un determinado delito y quién puede cometerlo, la policía puede intervenir de forma proactiva antes de que se produzca el presunto delito. Estas medidas suponen un cambio de una aplicación de la ley reactiva a una proactiva, cuyo objetivo es prevenir los delitos antes de que se cometan en lugar de detener reactivamente a las personas por delitos que ya han cometido.[66]

63. Egbert, Simon; Esposito, Elena y Heimstätt, Maximilian, *op. cit.*, pp. 189-200.
64. Ídem.
65. Ídem.
66. Ídem.

Pero esta forma de intervención pone en entredicho toda la separación estructural entre programas de decisión *condicionales/represivos* y de *propósito/preventivos* en la aplicación de la ley.

La desaparición de la distinción entre programación de *prevención/propósito* y programación de *represiva/condicional* tiene su mayor impacto en las predicciones policiales basadas en personas. La combinación de la programación de *prevención/propósito* y la programación de *represión/condicional* conduce a una situación en la que las predicciones algorítmicas se traducen directamente en medidas contra las personas que entrarían en la categoría de programas *represivos/condicionales*, que están por tanto mucho más estrictamente regulados y cuentan con fuertes salvaguardas.[67] En la actuación policial predictiva puede observarse, por lo general, que la prevención tiende a coincidir con la represión. Como consecuencia de la falta de separación entre ambos programas de decisión, se plantea la cuestión del tipo de programación empleada para orientar la actividad policial en su uso del programa predictivo y el tipo de restricciones que aplica.[68] Este contexto requiere un examen de las salvaguardas legales establecidas y de los mecanismos de control y balance como pilares clave del contrato social para la protección de las personas frente a un Estado excesivamente fuerte y arbitrario. La demarcación difusa entre los programas *condicionales* y de *propósito* en la aplicación de la ley provoca un giro hacia ideas de la aplicación preventiva de la ley y el concepto de "cultura de control" de Garlands.[69] De hecho, estos programas mixtos con predicciones informáticas pueden violar principios y conceptos rectores básicos de las acciones estatales como la proporcionalidad, la necesidad, la previsibilidad, etcétera.

El uso de programas informáticos de predicción incomprensibles y comprensibles intensifica esta demarcación difusa de la programación de decisiones de las fuerzas del orden. El *software* incomprensible en particular tiene un impacto inmenso en la reconfiguración de los programas de decisión en el sector policial. Al incorporar técnicas de aprendizaje automático, el *software* incomprensible utiliza *Big Data* para crear

67. Ibídem, p. 201.

68. Hildebrandt, Mireille, "New Animism in Policing: Re-animating the Rule of Law", en *The Sage Handbook of Global Policing*, Londres, 2016.

69. Garland, David, *Culture of Control: Crime and Social Order in Contemporary Society*, Oxford, Oxford University Press, 2001.

predicciones que no proporcionan explicaciones causales sobre la información subyacente a la toma de decisiones.[70] La lógica del *software* sigue el objetivo general del aprendizaje automático: "proporcionar precisión predictiva, incluso a expensas de la capacidad explicativa".[71]

Sin embargo, los objetivos se persiguen con medios destinados a contribuir causalmente a sus objetivos, por lo que el reconocimiento de las conexiones causales es esencial para los *programas de propósito*.[72] Así pues, es necesario identificar las interrelaciones causales entre medios y fines para limitar la total arbitrariedad de la programación y de las actividades de aplicación de la ley basadas en la informática.[73]

En la actualidad, la tecnología se está desarrollando tan rápidamente, cambiando las sociedades y la comunidad internacional a una velocidad incontrolable, que incluso los desarrolladores de nuevos modelos lingüísticos como Chat GPT4 están haciendo campaña por un mejor control de las herramientas de inteligencia artificial.[74] El enorme impacto y la evolución a través de nuevos modelos lingüísticos, que también son usados en el control del crimen, puede considerarse un salto cuántico, que lleva a los abordajes y conceptos de la aplicación de la ley de la era digital a la era cuántica. La construcción de cualquier número de contextos mediante el uso del conjunto infinito de *Big Data* crea una pluralidad de realidades. Como sabemos por la teoría cuántica, las realidades cambian en función de nuestra observación. Siguiendo el principio de incertidumbre, cuanto más exactamente conocemos la posición de una partícula en un sistema cuántico, menos conocemos su momento lineal.[75] El mismo principio podría aplicarse

70. Egbert, Simon; Esposito, Elena y Heimstätt, Maximilian, *op. cit.*

71. Ibídem, pp. 189-202; Esposito, Elena, "Transparency Versus Explanation: The Role of Ambiguity in Legal AI", en *Journal of Cross-disciplanary Reseach in Computational Law*, pp. 1-2. Disponible en: https://journalcrcl.org/crcl/article/view/10/8 [fecha de consulta 08/04/2024].

72. Luhmann, Niklas, *op. cit.*, p. 267 y ss.

73. Egbert, Simon; Esposito, Elena y Heimstätt, Maximilian, *op.cit*, pp. 189-202.

74. Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*. Disponible en: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [fecha de consulta 08/04/2024]

75. Hilgevoord, Jan y Uffink, Jos, "The Uncertainty Principle", en *The Stanford Encyclopedia of Philosophy*, 2023. Disponible en: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=qt-uncertainty. [fecha de consulta 08/04/2024].

también a otros pares de características de los sistemas cuánticos, como la energía y el tiempo. El siguiente chiste describe la situación de la mejor manera posible:

Werner Heisenberg conduce por la carretera cuando lo para un agente de tránsito. "Disculpe, señor", le dice el policía. "¿Sabe a qué velocidad iba?". "No", responde Heisenberg. "Pero sé exactamente dónde estoy".

El gato de Schrödinger es la descripción precisa de esta paradoja, porque el gato no observado está simultáneamente muerto y vivo hasta que se ve realmente que está vivo o muerto. El gato de Schrödinger es una buena metáfora para los modelos de control de la delincuencia y el uso de *software* predictivo, porque nunca podemos acercarnos a la realidad con una precisión arbitraria, sino sólo con afirmaciones estadísticas, centrándonos en una parte concreta. Por lo tanto, la predicción de una conducta probable desviada en la sociedad también se basa en una afirmación estadística sobre la aproximación del conocimiento humano, que, al convertir los pensamientos y la comunicación humanos en códigos e interacciones numéricas, distribuye en última instancia una pluralidad de posibilidades, cada una de las cuales puede realizarse de forma individual y/o simultánea. Este enfoque, sin embargo, no predice un concepto clave del contrato social: el Estado, como propietario del monopolio del poder, sólo puede utilizar este poder de forma proporcionada, transparente y previsible, por lo que las acciones se basan en decisiones comprensibles y entendibles. Los ciudadanos deben comprender las normas existentes y actuar y comportarse en consecuencia. En caso de una intervención de las autoridades estatales, esta lógica/fundamento para la intervención debe ser transparente y comprensible para los ciudadanos. De lo contrario, el sistema estatal se convierte en un sistema totalitario arbitrario en el que las acciones del Estado violan las ideas, normas y procedimientos manifestados en el contrato social: el individuo transfiere su derecho a usar la violencia/poder al Estado y, a cambio, el Estado, como propietario legítimo del monopolio del poder, debe garantizar el orden y la paz sociales mediante acciones transparentes, proporcionadas y basadas en el Estado de Derecho. [76]

76. Stanford Law School, "'Artificial Intelligence' Makes the Court System More Just, Efficient and Authoritative". Disponible en: https://law.stanford.edu/china-law-and-policy-association-clpa/articles/ [fecha de consulta 08/04/2024].

Por lo tanto, las medidas preventivas que puedan entrar en conflicto con los derechos y libertades individuales de los ciudadanos deben adoptarse de forma muy estricta, proporcionada y previsible. Cuanto menos intrusivas sean las acciones preventivas del Estado, menos reguladas deben estar. De ahí que se hayan desarrollado una pluralidad de "modelos preventivos" diferentes para regular las medidas estatales preventivas con el fin de mantenerlas dentro de los límites de una sociedad pacífica y previsible.[77] Uno de estos modelos se basa en un enfoque sanitario/médico, cuya lógica subyacente fue trasladada por Brantingham y Faust a la prevención de la delincuencia.[78]

Su modelo distingue tres niveles de prevención de la delincuencia:

1. prevención primaria, dirigida a modificar las condiciones criminógenas del entorno físico y social en general;
2. prevención secundaria, dirigida a la identificación e intervención tempranas en la vida de individuos o grupos en circunstancias criminógenas; y
3. Prevención terciaria, dirigida a la prevención de la reincidencia. [79]

La prevención primaria de la delincuencia identifica las condiciones del entorno físico y social que ofrecen oportunidades para la delincuencia. Este tipo de prevención pretende cambiar las condiciones hasta el punto de que no se cometa el delito. Se refiere a los esfuerzos para evitar que los problemas se produzcan en primer lugar.[80] Esta prevención *primaria de la delincuencia* es la situación de aplicación de la ley preventiva clásica/típica descrita anteriormente, en la que todavía existe una distinción muy clara entre programas *preventivos/represivos* o *de propósito/condicionales*. La probabilidad de que se produzca una violación de los derechos básicos y las libertades individuales es relativamente baja.

La prevención secundaria de la delincuencia se ocupa de la detección e identificación precoz de los delincuentes potenciales e intenta intervenir en sus actividades actuales de manera que no se cometan delitos.

77. Muir, Rick, "Taking prevention seriously: the case for a crime and harm prevention system", en *Strategic Review of Policing in England and Wales*, 2021, p. 9 y ss.

78. Brantingham, Paul J-, Faust, Frederic L., "A Conceptual Model of Crime Prevention", en *Crime & Delinquency*, 1976, pp. 284-296.

79. Brantingham, Paul J; Faust, Frederic L., *op. cit.*

80. Muir, Rick, *op. cit.*, pp. 1, 9.

La idea rectora es la intervención temprana, en una fase en la que un problema empieza a desarrollarse, lo que permite prevenirlo antes de que se manifieste.[81] En este caso, la intervención en las esferas protegidas de la libertad individual y el canon de los derechos básicos y humanos es ya mucho más fuerte; por lo que nos encontramos aquí a medio camino entre la distinción clásica de *programas de finalidad* y *condicionales*. La *prevención terciaria de la delincuencia se* ocupa de los delincuentes reales e interviene en sus esferas privadas para evitar que cometan más delitos. Aquí, la intervención preventiva se centra en la gestión de los problemas en curso, la evitación de futuras crisis y la reducción de las consecuencias perjudiciales.[82] Así pues, las medidas preventivas se inscriben claramente en el régimen especial de los derechos fundamentales y las libertades individuales. En consecuencia, la *prevención terciaria de la delincuencia* implica la consideración y aplicación de todo el canon de los derechos humanos y fundamentales. La intervención de las fuerzas policiales es tan fuerte que el individuo necesita la mejor protección posible frente a las probables intervenciones arbitrarias del Estado y su control.

Este modelo puede utilizarse como herramienta para analizar las medidas represivas predictivas basadas en la informática. Una demarcación clara entre medidas *preventivas/represivas* o programas *de propósito/condicionales* se hace casi imposible, lo que conduce a una elusión de las normas mínimas de protección establecidas por el contrato social y manifestadas por el Estado de Derecho general. Las tres categorías mencionadas podrían servir de orientación para establecer una nueva matriz de categorización de un régimen de protección y regulación de las medidas predictivas de las fuerzas y cuerpos de seguridad. Esta herramienta de evaluación y control podría ser la futura pauta para la aplicación de la inteligencia artificial y otras herramientas informáticas en la actuación policial predictiva. La división en tres niveles de prevención puede utilizarse para categorizar las respectivas medidas y aplicar en consecuencia las normas mínimas para los intereses jurídicos protegidos.

81. Ídem.
82. Ídem.

Si no regulamos y limitamos la participación de los sistemas autónomos de toma de decisiones en el sistema de control social y aplicación de la ley, permitiremos que algoritmos de aprendizaje autónomo intervengan en la comunicación social, dando lugar a interacciones basadas en predicciones generadas por computadora. Esto elevaría la idea general del contrato social y el Estado de Derecho a una nueva esfera: El contrato social entre los ciudadanos y el Estado se convertiría en un contrato entre individuos y programas informáticos y el Estado de Derecho en un Estado de algoritmos.

Conciliar la Inteligencia Artificial y la humana. Complementar y no suplantar la sentencia judicial

Mathis Schwarze y Julian Roberts*

¿Cómo podría el aprendizaje automático mejorar la toma de decisiones judiciales, o el proceso de imposición de penas en general? Las propuestas para incorporar la IA al proceso de imposición de penas van de lo modesto a lo ambicioso, desde el mero apoyo a los jueces hasta su sustitución total. La forma más plausible para la contribución de la IA es una sentencia consultiva para guiar a los jueces. Dado que el algoritmo aprende de los juicios humanos, nos referimos a esto como Aprendizaje Algorítmico Humano o HAL (Human-Algorithmic Learning) para abreviar. El HAL también puede enriquecer las sentencias mediante contribuciones a la orientación proporcionada por las Cámaras de Apelación y, en un número creciente de jurisdicciones, los lineamientos de sentencia.

Introducción

Este ensayo comienza apuntando brevemente algunas razones por las que una sentencia producto de un aprendizaje automático representa una alternativa poco convincente a la toma judicial de decisiones. Esta sección aborda la característica central de la imposición de penas en los países de *common law*: la audiencia de sentencia, que constituye una fase distintiva del proceso penal. Según muchas propuestas, la audiencia de sentencia se volvería innecesaria en la mayoría

* Profesor de la Universidad de Oxford y director ejecutivo de la *Sentencing Academy*. En 2022 fue nombrado *King's Counsel* (Honoris Causa) por sus contribuciones en materia de sentencias internacionales. En 2021 fue galardonado con el Premio Sellin-Glueck de la *American Society of Criminology* por su investigación sobre la justicia penal de manera internacional y comparativa.

de los casos. La segunda sección explora algunos modos en los que la IA podría complementar la toma de decisiones judiciales. Los sistemas informáticos asistidos por computadora son el enfoque más debatido y ya se han aplicado varios desde al menos la década de 1980. En esta sección se analizan los problemas que plantea la construcción de una base de datos adecuada y las ventajas de una *sentencia consultiva* derivada de la IA. En la tercera parte se analizan brevemente otras formas en que la IA podría mejorar el proceso de imposición de penas, por ejemplo, detectando fuentes de sesgo y desviaciones de los principios para la imposición de penas más eficazmente que los medios convencionales. También señalamos formas inexploradas en que la IA podría mejorar los lineamientos proporcionados a los tribunales.[1]

Por qué la IA debe complementar a los jueces, no sustituirlos

Bagaric y Wolf[2] ofrecen una evaluación optimista de las sentencias informatizadas, argumentando que el proceso es "extremadamente adaptable a una toma de decisiones informatizada"[3]. Concluyen que "la sentencia informatizada es preferible a la sentencia judicial".[4] En nuestra opinión, la imposición de penas es menos adaptable a la computarización de lo que sugieren, y es improbable que la IA sustituya la toma de decisiones judiciales en la imposición de penas, por razones tanto tecnológicas como deontológicas. Las dificultades tecnológicas de programar un "juez informático" surgen de la naturaleza del razonamiento judicial, y de los desafíos para introducir los datos relevantes para decidir la condena.

El razonamiento jurídico humano no es un proceso puramente lógico. Algunas cuestiones jurídicas pueden resolverse utilizando "reglas

1. Ryberg, Jesper; Roberts, Julian, *Sentencing and Artificial Intelligence*, New York, Oxford University Press, 2022.

2. Bagaric, Mirko; Wolf, Gabrielle, "Sentencing by Computer: Enhancing Sentencing Transparency and Predictability, and (Possibly) Bridging the Gap between Sentencing Knowledge and Practice", en *George Mason Law Review*, 2018, p. 681.

3. Stobbs, Nigel; Hunter, Dan; Bagaric, Mirko, "Can Sentencing Be Enhanced by the Use of Artificial Intelligence?", en *Criminal Law Journal*, 2017, p. 272.

4. Bagaric, Mirko; Wolf, Gabrielle, *op. cit.*, p. 708.

de manual" relativamente lógicas. Sin embargo, los jueces se enfrentan a veces a "casos difíciles" para los que las normas existentes son insuficientes. Estos casos requieren la interpretación de una norma o la creación de nuevas normas con referencia a "objetivos de política pública o a los requisitos de la justicia".[5] Esta tarea requiere cualidades fundamentalmente *humanas*, como el sentido común y la conciencia moral, social y cultural.[6] Por su naturaleza, la condena de un delincuente es un ejemplo de "caso difícil". Funciona sobre la base de la discreción judicial y la individualización y tiene un elemento moral. Por lo tanto, sustituir a los jueces sentenciadores por ordenadores requeriría una IA de "nivel humano" o una "inteligencia general artificial" –el "santo grial" en este campo– que sigue siendo difícil de alcanzar.[7]

Tampoco está claro cómo la IA podría ayudar a aclarar la relevancia normativa de los elementos del proceso de imposición de penas. Supongamos una jurisdicción en la que el poder legislativo ha prescrito que las sentencias deben ajustarse al principio de proporcionalidad retributiva. El reto para los tribunales es aplicar este principio, por ejemplo, teniendo en cuenta factores relacionados con la proporcionalidad y descartando o restando peso a otros factores. En un régimen de este tipo, el nivel de daño infligido tendría mucho peso, mientras que el riesgo de reincidencia del delincuente apenas tendría influencia en la condena. A continuación, los jueces deben decidir cuáles de los múltiples factores de la sentencia son relevantes para ese principio. ¿Es relevante el remordimiento? ¿La premeditación debe tener mucho o poco peso? En la actualidad, los jueces toman estas decisiones individualmente. Es difícil concebir cómo podría programarse la IA para diferenciar los factores en función de su relevancia para una sentencia proporcional. El aprendizaje automático podría ser útil, sin embargo, para determinar si los tribunales se ajustan de hecho al principio de proporcionalidad.[8]

5. Dworkin, Ronald, "Judicial Discretion", en *The Journal of Philosophy*, 1963, p. 628.

6. Susskind, Richard, "Detmold's Refutation of Positivism and the Computer Judge", en *The Modern Law Review 49*, 1986, p. 133.

7. Boden, Margaret, *AI: Its Nature and Future*, Oxford, Oxford University Press, 2016, p. 21.

8. Chiao, Vincent. "Predicting Proportionality: The Case for Algorithmic Sentencing". *Criminal Justice Ethics*, 2018, pp. 238-261.

Las preocupaciones éticas tienen su origen en la inaptitud de los ordenadores para cualquier "función humana que implique respeto interpersonal, comprensión y amor".[9] Una comparación entre la aplicación de la IA al derecho y a la medicina lo ilustra. En el futuro, la mayoría de los diagnósticos se realizarán a distancia mediante herramientas de diagnóstico de IA y es probable que el movimiento hacia la "telemedicina" se haya acelerado por la pandemia de 2020. No hay que subestimar el papel de la empatía en la medicina (por ejemplo, cuando hay que comunicar un diagnóstico con consecuencias graves). Sin embargo, la principal preocupación de los pacientes es la exactitud del diagnóstico y la eficacia del tratamiento posterior. A diferencia de la medicina, la sentencia tiene una dimensión normativa. Y lo que es más importante, un ordenador seguramente sería menos eficaz a la hora de comunicar una censura, ya que carece de un compromiso moral con su resultado. Es probable que recibir una carta u otra forma de comunicación tenga menos efecto censurador que una declaración de reprobación, expresada por una autoridad legítima que actúa en un foro público.

¿Pérdida de audiencia?

Como el término indica, las audiencias donde se da lectura del veredicto (o en jurisdicciones sin una audiencia separada, los alegatos) dan a las partes la oportunidad de exponer sus puntos de vista, perspectivas y recomendaciones al juez. En esta fase del proceso, tanto los imputados como las víctimas buscan una *audiencia* justa. Es cierto que, en la actualidad, las lecturas de veredicto pueden ser asuntos superficiales en los que el tribunal impone la condena a un imputado silencioso cuyo abogado se comunica en su nombre. Esta realidad no debe hacernos perder de vista los potenciales beneficios derivados de un debate exhaustivo de las cuestiones implicadas. El proceso de imposición de penas, a menudo pasado por alto en la literatura del campo, requiere cierta consideración a medida que prolifera el uso de la HAL (y otros avances tecnológicos como las audiencias en línea). Los defensores de la IA dicen poco sobre el procedimiento a seguir cuando

9. Weizenbaum, Joseph, *Computer Power and Human Reason: From Judgment to Calculation*, San Francisco, 1976, p. 269.

el programa ha generado una sentencia. ¿Aparece como archivo adjunto a un correo electrónico dirigido al imputado?

Si HAL es más eficaz que los jueces humanos a la hora de determinar la sentencia más apropiada –ya se refiera esto a principios retributivos o preventivos–, es difícil pensar que haya lugar para una audiencia presencial. Suponiendo que toda la información pertinente fue introducida en el programa, ¿qué queda por debatir? No parece haber papel ni oportunidad para que los imputados expresen su punto de vista. El derecho a ser oído es una garantía procesal importante. Los delincuentes siempre tienen derecho a ser oídos, y los abogados defensores hacen un pedido de mitigación en la sentencia. Estas presentaciones pueden contener información sustancial no disponible hasta ese momento, o la interpretación de los hechos puede influenciar en la decisión de un juez humano. Este elemento de la sentencia parece quedar fuera de juego cuando un programa informático decide la sentencia. De Mulder y Gubby ya lo señalaron hace casi 40 años: "Si no hay ningún agente humano disponible a quien expresar su punto de vista, su propósito se verá frustrado. Semejante frustración del propósito podría fácilmente afectar a la legitimación de la autoridad".[10]

Se podría aplicar algún tipo de opción de "hablar antes de la sentencia". El imputado podría tener la oportunidad de presentar una declaración previa a la sentencia final de HAL. Podría consistir en una coartada, arrepentimiento, disculpa o explicación del delito. En consonancia con el sistema acusatorio, podría ser objeto de comentarios por parte de la fiscalía. O un juez podría revisar la declaración y, si está convencido de la necesidad de intervenir, podría introducir una "corrección por atenuación personal" que modificaría la sentencia final de HAL. Pero desde la perspectiva de los actores, esto es un pobre sustituto de la oportunidad del imputado de dirigirse al juez en audiencia pública.

Sin audiencia, otros participantes también podrían perder la voz. Aunque son pocas las audiencias en las que las víctimas leen sus declaraciones en voz alta públicamente, hay pruebas de que, cuando esto ocurre, la experiencia es positiva para las víctimas.[11] Pueden tener una

10. De Mulder, Richard; Gubby, Helen, "Legal Decision Making by Computer: An Experiment with Sentencing", en *Computer Law Journal*, 1983, p. 302.

11. Roberts, Julian, "Listening to the Crime Victim: Evaluating Victim Input at Sentencing and Parole", en *Crime and Justice*, Chicago, 2009.

sensación de reivindicación o participación. Este no sería el caso si simplemente presentaran su Declaración de Repercusiones para la Víctima (VIS) a través de un portal en línea, o peor aún, si se limitaran a responder a una serie de preguntas (por ejemplo, "¿Sufrió daños físicos? En caso negativo, pase directamente a la P8"). Por último, existen pruebas de que escuchar a la propia víctima hablar del impacto del delito tiene un efecto más poderoso sobre los delincuentes. Cuando el VIS simplemente forma parte de la presentación de la acusación, el delincuente puede estar menos atento. Esta (y otras) funciones comunicativas de la lectura de la sentencia –o de los alegatos– se pierden cuando HAL está en el estrado. Los imputados a veces se quejan de que sus puntos de vista no son escuchados por los jueces en un tribunal muy ocupado. Pero si los jueces son duros de oído, HAL es sordo como una tapia.

El problema de los datos ingresados

El principal obstáculo para crear un algoritmo de imposición de penas que funcione se encuentra en la fase de introducción de datos. Los defensores de la IA han pasado por alto la complejidad de introducir la información adecuada. Para construir una base de datos de sentencias que sirva de base a la IA, es necesario que las sentencias se introduzcan de forma habitual. Construir una base de datos para una herramienta de diagnóstico médico es relativamente sencillo. Los pacientes responden a una serie de preguntas: ¿el dolor es agudo o sordo? ¿Cuál es la intensidad del dolor en una escala de 10 puntos? Una vez que muchos pacientes han introducido sus respuestas, el programa determina los correlatos y predictores de síntomas y patologías. En lo subsiguiente, basándose en los resultados de terapias anteriores, el programa recomienda un tratamiento específico.

La sentencia es cualitativamente diferente. Un tribunal tiene en cuenta muchos factores sutiles e interconectados cuyo peso varía en función del delito y del delincuente. Stobbs *et al* afirma que "los únicos datos que habría que introducir (en el programa) en un caso concreto son los factores agravantes y atenuantes que fueran pertinentes".[12] Esto sugiere que el proceso es tan sencillo como introducir el número, la naturaleza

12. Stobbs, Nigel; Hunter, Dan; Bagaric, Mirko, "Can Sentencing Be Enhanced by the Use of Artificial Intelligence?", en *Criminal Law Journal*, 2017, p. 272.

y la duración de los síntomas en un programa de diagnóstico. La imposición de penas es más complicada; los factores tienen pesos diferentes en contextos diferentes, y su importancia no puede determinarse de antemano ni reducirse necesariamente a un valor específico (por ejemplo, remordimiento: "alto"; antecedentes penales: "modestos").

Algunas características de los casos pueden prestarse a ser introducidos a la base de datos. Por ejemplo, el sexo o la edad del delincuente (suponiendo que se consideren relevantes para la sentencia). Otras variables son tan claras como las médicas: ¿se declaró culpable el delincuente? ¿Se declaró culpable en una fase temprana o tardía del procedimiento? ¿Tiene condenas previas relevantes? En caso afirmativo, ¿cuántas? Las condenas previas ilustran las ventajas y las limitaciones de las sentencias asistidas por IA. Los antecedentes penales de un delincuente podrían reducirse a un cómputo mecánico, pero, una vez más, la IA sólo sería capaz de ponderar crudamente los delitos anteriores. ¿Qué relevancia tiene una condena previa por robo para la condena actual por homicidio? Si dos delincuentes tienen la misma condena previa por agresión, pero una ocurrió hace dos años y la otra hace diez, ¿debería, y hasta qué punto, el tribunal sopesar la antigüedad de la pena? La IA podría aplicar reglas sencillas, como asignar un incremento de gravedad a cada delito anterior, pero que disminuya con el tiempo. Pero el juicio humano seguirá siendo necesario. La IA puede recomendar un aumento de la condena de un año cuando el delincuente tenga una condena anterior grave por el mismo delito que está siendo procesado. Los detalles legalmente significativos probablemente escapen al análisis de la IA. Si el delito anterior se produjo como resultado de alguna presión situacional que reapareció en la vida del delincuente en el momento del delito actual, un responsable humano puede inclinarse por descartar el incremento por antecedentes o ignorar por completo la condena anterior.

Se pasarían por alto muchas otras consideraciones cualitativas que tendrían un efecto significativo: el delincuente profundamente arrepentido frente al delincuente que expresa superficialmente su arrepentimiento a través de su abogado; el que tiene condenas previas pero que también una explicación convincente para esos delitos. Puede introducirse en la base de datos una gran cantidad de información sobre los factores de la condena, pero es difícil ver cómo pueden tenerse en cuen-

ta la interpretación, el contexto y la explicación, proporcionados por el abogado defensor en el alegato (véase más adelante).

Quienes impartan sentencia podrían introducir los datos cuando condenan a un delincuente, indicando de forma más cualitativa la importancia asignada a los distintos factores. Esto llevaría a una predicción más precisa, pero exigiría mucho trabajo a los jueces. Ese es un problema en sí mismo. Uno de los retos de las bases de datos existentes es la fatiga jurídica a la hora de introducir datos.

En resumen, las sentencias recomendadas por IA sólo se aproximarían a la sentencia "verdadera", que refleja todas las características relevantes del caso, porque los datos en los que se basa captan sólo de un modo imperfecto los componentes que determinaron las sentencias anteriores. En general, por lo tanto, *sustituir* al juez por la IA no es una opción realista ni saludable. Esto no quiere decir que un "juez informático" no sea concebible en asuntos jurídicos relativamente simples e intrascendentes –por ejemplo, Estonia está explorando el uso de la IA para decidir sobre pequeñas demandas civiles–[13], pero la sentencia no es uno de ellos.

La sentencia consultiva

¿Cómo podría entonces la IA ayudar a los jueces a dictar sentencia? HAL podría seguir los pasos de los "sistemas informatizados de sentencia" pioneros que se probaron en los años ochenta y noventa,[14] con la principal diferencia del enorme aumento de la potencia informática disponible en la actualidad. Estos sistemas se desarrollaron porque faltaba un método "para que cualquiera [...] pueda saber de forma sistemática, actualizada y accesible, y de manera constante, qué tipo de sentencias se están dictando".[15] La aparición de los linea-

13. Dymitruk, Maria, "Artificial Intelligence as a Tool to Improve the Administration of Justice?", en *Acta Universitatis Sapientiae Legal Studies*, 2019, pp. 179-190.

14. Doob, Anthony; Park, Norman W., "Computerized Sentencing Information for Judges: An Aid to the Sentencing Process", en *Criminal Law Quarterly*, 1987, pp. 54-72; Simon, Eric; Gaes, Gerry, "ASSYST - Computer Support for Guideline Sentencing", en *Proceedings of the Second International Conference on Artificial Intelligence and Law - ICAIL '89*, Vancouver, CM Press, 1989, pp. 195-200.

15. The Canadian Sentencing Commission 1987, párr. 60.

mientos de imposición de penas mejoró la información judicial sobre la práctica de imposición de penas que reflejaban estos instrumentos, pero la profundidad de la información es inevitablemente limitada. HAL llenaría este vacío basándose en la práctica anterior para proporcionar una sentencia para un caso individual, que denominamos *Sentencia Consultiva (SC*). Como HAL funcionaría basándose en la práctica anterior, la SC podría incorporar consideraciones preventivas y orientadas a la proporcionalidad, en función del marco de imposición de penas pertinente.[16]

La SC de HAL se deriva de clasificadores de aprendizaje automático. Este enfoque tecnológico parece preferible a la codificación manual de reglas fijas en el programa.[17] Aunque este enfoque basado en la lógica proporcionaría un alto grado de transparencia (con reglas tales como "si el delincuente tiene x condenas previas, *entonces…*" que son fáciles de entender para los humanos), un método basado en ejemplos promete resultados mucho más refinados y precisos.[18] Además, para los fines aquí previstos, una capacidad de aprendizaje automático es esencial porque permitiría hacer un seguimiento de la práctica habitual de imposición de penas.

Los jueces podrían correr el programa HAL en paralelo a la consulta de los lineamientos para la imposición de penas pertinentes. Para evitar proponer una sentencia precisa y única, el programa debería ofrecer intervalos de confianza para sus recomendaciones. Por ejemplo, el programa podría arrojar una pena de 18 meses, con un intervalo de confianza del 95% que iría de 15 a 25 meses y un punto medio de 20 meses, que comprende el 95% de las penas privativas de libertad impuestas por este delito.

La SC serviría como punto de partida para considerar las circunstancias del caso que se presenta para sentencia, junto con los alegatos

16. Hutton, Neil, "Sentencing, Rationality, and Computer Technology", en *Journal of Law and Society*, 1995, pp. 549-570; Chiao, Vincent, "Predicting Proportionality: The Case for Algorithmic Sentencing", *Criminal Justice Ethics*, 2018, pp. 238-261.

17. Bagaric, Mirko; Wolf, Gabrielle, "Sentencing by Computer: Enhancing Sentencing Transparency and Predictability, and Bridging the Gap between Sentencing Knowledge and Practice", en *George Mason Law Review*, 2018.

18. Chiao, Vincent, "Sentencing and the right to reasons"; Ryberg, Jesper, "Sentencing and Algorithmic Transparency", en *Sentencing and artificial intelligence*, Nueva York, Oxford University Press, 2022.

de los abogados, los dictámenes, las declaraciones de las víctimas y otros orientadores como las resoluciones de las apelaciones. El sistema estaría a disposición del público y sería transparente, conforme a los requisitos expuestos por varios colaboradores de este volumen. Las partes utilizarían la SC para preparar sus proyectos de sentencia. Los ciudadanos podrían introducir información y recibir una SC sobre los datos que proporcionen, tal como las personas introducen los datos de un conductor y de un vehículo en un sitio web y obtienen una estimación de las primas del seguro del vehículo.

Aspectos problemáticos del uso de la práctica previa como fuente

Si los tribunales utilizaran la SC, las sentencias serían más coherentes y transparentes (véase más adelante). Sin embargo, utilizar la práctica judicial previa como marco de referencia también tiene sus aspectos problemáticos. Un desafío para alimentar la base de datos del programa con sentencias anteriores es garantizar que las respuestas tengan un fundamento basado en principios. Este es el caso en sistemas de imposición de penas que afirman fundamentos tanto retributivos como preventivos. Por ejemplo, tras los disturbios en Inglaterra en 2011, los tribunales se desviaron del castigo proporcionado para fomentar la disuasión generalizada. Este enfoque fue respaldado por el Tribunal de Apelación, que declaró que "la imposición de penas severas, destinadas a proveer tanto castigo como disuasión, debe proseguir".[19]

Si estos juicios fueran introducidos en la base de datos del programa, su predicción para una pena proporcional "estándar" se distorsionaría. Esto comprometería los principios en la predicción del algoritmo, y socavaría así la función comunicativa de la pena (por ejemplo, Duff).[20] Una solución sería que el marco de imposición de penas estuviera integrado al algoritmo. Podría afirmar la proporcionalidad retributiva como razonamiento principal, o la prevención. Si la proporcionalidad retributiva fuera considerada primordial, las desviaciones del castigo proporcionado por razones de disuasión, incapaci-

19. EWCA, Crim Nº 2312, "Blackshaw, R y otros", 2011, párr. 4.

20. Duff, Antony, *Punishment, Communication, and Community*, Oxford, Oxford University Press, 2001.

tación y rehabilitación se harían evidentes. De este modo, el programa podría predecir la proporcionalidad "pura" pero indicar dónde podrían justificarse las desviaciones.

Otro inconveniente de reflejar la práctica actual de imposición de penas en las predicciones del programa es que se institucionalizaría cualquier sesgo existente. Por ejemplo, se supone que una declaración de culpabilidad reduce la condena independientemente de la naturaleza o gravedad del delito. Sin embargo, algunos jueces tienden a "restar importancia" a la declaración en los casos más graves, en contra de los lineamientos publicados. Este mal uso sería captado por la base de datos y luego reforzado a través de las recomendaciones de la IA u otros tipos de resultados. Además, es posible que la arquitectura (y los resultados) del algoritmo cambien con el tiempo de forma tal que la autoridad a cargo de los lineamientos no los apruebe. Por ejemplo, un factor agravante o atenuante específico enumerado en los lineamientos podría ser ignorado en gran medida por los tribunales, por lo que no tendría ningún impacto en la SC emitida por el programa. Si los jueces se guían únicamente por la SC, esto significaría que son los tribunales de primera instancia quienes determinan la práctica de imposición de penas, en lugar de la autoridad competente en materia de lineamientos o el tribunal de apelación. Otras partes interesadas en el proceso penal, como los representantes de las víctimas, los abogados o los académicos, que actualmente participan en la elaboración de los lineamientos, perderían toda influencia en la imposición de penas.

Por estas razones, es necesaria una institución que asuma una función correctiva. Esta función sería competencia de las actuales autoridades encargadas de los lineamientos para la imposición de penas. Al emitir lineamientos sobre la imposición de penas y configurar así la práctica judicial, el Consejo o la Comisión de Sentencias influiría *indirectamente* en la base de datos del algoritmo. Esto le permitiría contrarrestar cualquier sesgo o evolución injustificada que se pusiera de manifiesto en el programa. Con el tiempo, el programa y los lineamientos de imposición de penas se fusionarían en un único sistema. El Consejo de Sentencias tendría autoridad para ajustar directamente el algoritmo. Esto convertiría la SC en una combinación tanto de una predicción de la IA basada en la práctica anterior como de intervenciones normativas realizadas por el consejo de lineamientos para la

imposición de penas. En cualquier caso, las comisiones de sentencia no deberían ceder competencias al algoritmo.

Ventajas del sistema de sentencia consultiva: promoción de una mayor uniformidad

Aunque la HAL no pueda (o no deba) sustituir a los jueces a la hora de dictar sentencia, es capaz de contribuir al proceso de imposición de penas mucho más de lo que lo hace en la actualidad. Más allá del limitado (y controvertido) papel de la predicción del riesgo, en la actualidad la IA no contribuye al proceso de imposición de penas. En las secciones restantes de este ensayo identificamos otras formas en las que la IA puede mejorar las decisiones de imposición de penas. Comenzamos por el objetivo más obvio: la variación injustificada en los resultados de las sentencias.

Las investigaciones realizadas durante más de un siglo han documentado disparidades en las sentencias debidas a la personalidad y las actitudes del juez. Los primeros intentos de utilizar computadoras en la imposición de penas pretendían reducir la disparidad proporcionando a los jueces recomendaciones basadas en las características clave del caso y en otras informaciones. Durante mucho tiempo, el objetivo principal de la reforma de la imposición de penas ha sido reducir las disparidades y la discriminación en las sentencias. Esto es especialmente cierto en el caso de los lineamientos para la imposición de penas.[21] En Estados Unidos e Inglaterra y Gales (y en muchas otras jurisdicciones), se introdujeron lineamientos para la imposición de penas para estructurar las decisiones judiciales.[22] La investigación sugiere que estas reformas han logrado una mayor coherencia en la imposición de penas,[23] aunque la

21. Crackanthorpe, Montague, "Can Sentences be Standardised?", en *The Nineteenth Century*, 1900, pp. 103-115; Frankel, Marvin E., "Lawlessness in Sentencing", *University of Cincinnati Law Review*, 1972, pp. 1-54.

22. Frase, Richard, "Forty Years of American Sentencing Guidelines: What Have We Learned?", *Crime & Justice*, 2019, pp. 79-135; Roberts, Julian; Ashworth, Andrew, "The Evolution of Sentencing Policy and Practice in England and Wales, 2003-2015", en *Crime & Justice*, 2016, pp. 307-358.

23. Pina-Sánchez, Jose; Linacre, Robin, "Enhancing Consistency in Sentencing: Exploring the Effects of Guidelines in England and Wales", en *Journal of Quantitative Criminology*, 2014, pp. 731-748.

preocupación por la disparidad y la discriminación persisten.[24] ¿Podrá HAL cumplir esta tarea con mayor eficacia? Su potencial para reflejar combinaciones complejas de criterios es una clara ventaja sobre los lineamientos para la imposición de penas, que solo pueden funcionar basándose en una lógica simple de cuadrícula o diagrama de flujo.

Utilizando la SC, los jueces se beneficiarían de una fuente de información imparcial sobre el impacto que los factores han tenido en los resultados de las sentencias. Esta información puede influenciar sus decisiones posteriores. En la actualidad, la orientación es episódica (procedente de decisiones de apelación) o está potencialmente distorsionada por sesgos acusatorios. En cuanto a lo primero, algunos lineamientos indicarán el peso que debe tener un factor; en cuanto a lo segundo, el abogado puede señalar la jurisprudencia que indicaría el impacto que un factor debe tener en la sentencia. La calibración de la IA proporcionaría una guía mucho más precisa e imparcial sobre el peso que los tribunales han asignado a los factores para la condena.

La coherencia aumentaría probablemente a medida que los tribunales reflejen la totalidad de las prácticas actuales en sus sentencias. El algoritmo arrojaría la misma condena independientemente de dónde, cuándo y por quién fuera utilizado, siempre que se introduzcan en él el mismo conjunto de datos. Habría una tendencia a que las sentencias fueran menos dispares en torno a la sentencia promedio, ya que los jueces que vieran que una sentencia está fuera de la distribución normal probablemente la aproximarían más a la sentencia media. Sin embargo, la transferencia entre las impresiones de los jueces durante el juicio y lo que de hecho se introduce en el algoritmo sigue siendo una puerta de entrada para las incoherencias. De forma similar al efecto de los lineamientos de imposición de penas, los jueces podrán "generar" un resultado de predicción deseado si introducen los criterios adecuados. A continuación, se destacan algunas áreas específicas en las que la IA podría aumentar la coherencia.

24. Pina-Sánchez, Jose; Robin Linacre, "Sentence Consistency in England and Wales: Evidence from the Crown Court Sentencing Survey", en *British Journal of Criminology*, 2013, pp. 1118-1138.

Reducir e identificar la discriminación directa e indirecta

Uno de los sesgos cognitivos más preocupantes está relacionado con características legalmente protegidas como la raza y el género. Las disparidades raciales en la justicia penal son un fenómeno presente en la mayoría de las jurisdicciones occidentales.[25] Sin embargo, no hay conocimiento suficiente sobre el alcance y las causas de estas disparidades, en parte debido a la falta de datos adecuados sobre las sentencias.[26] La investigación sobre el sesgo racial sugiere que los jueces "piensan en la raza como una heurística relevante y útil para determinar la culpabilidad del imputado y la perniciosidad del delito".[27]

La IA podría ayudar a reducir la discriminación en las sentencias, una función, sin embargo, que dista de ser sencilla. Como señalan Bagaric y Wolf:

> … las computadoras no tienen prejuicios instintivos e inconscientes, son incapaces de discriminar inadvertidamente y no se ven influidos por consideraciones ajenas ni por suposiciones y generalizaciones que no estén integradas en sus programas.[28]

Así pues, excluir la raza y otras características protegidas de los delincuentes como variables de las predicciones del algoritmo parece prometer recomendaciones de penas imparciales. Sin embargo, emplear una herramienta de aprendizaje automático conlleva el riesgo específico de generar un "sesgo algorítmico". Si hubiera discriminación en la base de datos del HAL, la simple exclusión de los factores protegidos sería insuficiente para garantizar predicciones imparciales. El algoritmo aprendería a sustituir los factores excluidos por "variables sustitutivas". Por lo tanto, la introducción del HAL tendría que ir acompañada

25. Phillips, Coretta; Bowling, Ben, "Ethnicities, Racism, Crime, and Criminal Justice", en *The Oxford Handbook of Criminology*, Oxford, Oxford University Press, 2017, pp. 190-212.

26. Pina-Sánchez, Jose; Roberts, Julian; Sferopoulos, Dimitrios, "Does the Crown Court Discriminate Against Muslim-named Offenders? A Novel Investigation Based on Text Mining Technique", en *British Journal of Criminology*, 2019, pp. 718-736.

27. Eberhardt, Jennifer; Davies, Paul; Purdie-Vaughns, Valerie; Johnson, Sheri Lynn, "Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes", en *Psychological Science*, 2006, pp. 383-386.

28. Bagaric, Mirko; Wolf, Gabrielle, "Sentencing by Computer: Enhancing Sentencing Transparency and Predictability, and Bridging the Gap between Sentencing Knowledge and Practice", en *George Mason Law Review*, 2018, p. 696.

de una función de investigación que examinara incluso los factores aparentemente neutros en busca de un impacto desproporcionado. Sin embargo, la forma más exhaustiva de confrontar las predicciones sesgadas sería abordar directamente las fuentes de discriminación que influyen en los datos de entrenamiento del algoritmo.

La IA podría contribuir a esta tarea identificando más eficazmente las fuentes de discriminación directa e indirecta en la imposición de penas. El enfoque actual para detectar la discriminación *directa* implica la realización de análisis de variables múltiples que comparen los resultados de las sentencias de, por ejemplo, imputados de minorías visibles y blancos. Tras controlar el mayor número posible de características jurídicas, ¿hasta qué punto difieren las pautas de imposición de penas de los grupos raciales? En la mayoría de los países occidentales se han publicado análisis de este tipo.[29] Un algoritmo de imposición de penas podría facilitar enormemente esta investigación. Suponiendo que funcionara sobre la base de una amplia gama de variables legales, podría mostrar disparidades raciales simplemente manipulando la variable "raza" y realizar un seguimiento de los cambios en la disparidad racial a lo largo del tiempo. La discriminación *indirecta* derivada de algún elemento del régimen de imposición de penas es más difícil de detectar para los investigadores humanos. Para la IA, sería relativamente sencillo. La declaración de culpabilidad y los antecedentes penales del imputado son dos ejemplos de fuentes indirectas de impactos desproporcionados.

Las reducciones de condena basadas en la declaración de culpabilidad pueden discriminar indirectamente a las minorías.[30] Los imputados negros tienen menos probabilidades de declararse culpables, por una serie de razones, entre ellas la "sobreacusación" por parte de los fiscales, la falta de confianza en el *Corpus Juris Secundum* o un

29. Sentencing Council of England and Wales, "Investigating the association between an offender's sex and ethnicity and the sentence imposed at the Crown Court for drug offences", 2019. Disponible en: https://www.sentencingcouncil.org.uk/wp-content/uploads/Sex-and-ethnicity-analysis-final-1.pdf [fecha de consulta: 12/04/2024]; Hood, Roger, Race and Sentencing: a study in the Crown Court: a report for the Commission for Racial Equality, Oxford, Oxford University Press, 1992; Roberts, Julian; Doob, Anthony, "Race, Ethnicity and Criminal Justice in Canada", en *Crime & Justice*, 1997, pp. 469-522.

30. Johnson, Brian; Richardson, Rebecca, "Race and Plea Bargaining", en *A System of Pleas*, Nueva York, Oxford University Press, 2019.

asesoramiento jurídico inadecuado.[31] La consecuencia es que estos imputados tienen menos probabilidades de beneficiarse de las reducciones de penas basadas en la declaración de culpabilidad: es más probable que sean encarcelados, y por un período más largo. Por este motivo, varios autores han pedido la abolición de las reducciones basadas en la declaración de culpabilidad.[32] Un argumento similar se ha esgrimido contra los aumentos de antecedentes penales, que se traducen en tasas de detención más elevadas y condenas más largas para las minorías raciales. El resultado es la desproporcionalidad racial en las poblaciones penitenciarias.[33] Es probable que haya otros aspectos ostensiblemente neutrales de la imposición de penas que desencadenen una discriminación indirecta, y la IA sería mucho más eficaz para descubrir estos efectos.

Detección de anomalías locales en la imposición de penas y desviaciones de los principios

Hasta ahora nos hemos centrado en la sentencia, no en el juez. Suponiendo una base de datos relativamente completa que incluya las características de los casos y los motivos de la sentencia, la IA podría centrarse en el juez. La IA serviría para identificar tribunales aberrantes o jueces deshonestos, en términos de la práctica actual. Si la IA revelara indicios de disparidad regional en la imposición de penas, permitiría al consejo de imposición de penas identificar a estos tribunales o jueces, a fin de permitir la adopción de medidas correctoras.

Algunos tribunales locales pueden desarrollar sus propias prácticas desproporcionadas. Por ejemplo, la disparidad regional de las penas está muy extendida en Alemania: Un estudio reciente halló diferencias regionales en la duración media de las penas de hasta un 25 % en todos

31. Lammy, David, *An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System.* HMSO. Disponible en: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643001/lammy-review-final-report.pdf [fecha de consulta: 12/04/2024]; Hood, Roger, *Race and Sentencing: a study in the Crown Court: a report for the Commission for Racial Equality*, Oxford, Oxford University Press, 1992.

32. Tonry, Michael, "Abandoning Sentence Discounts for Guilty Pleas", en *In Principled Sentencing. Readings on Theory and Policy*, Oxford, 2009.

33. Frase, Richard S; Roberts, Julian, *Paying for the Past: Prior Record Enhancements in the US Sentencing Guidelines*, Nueva York, Oxford University Press, 2019.

los delitos al comparar las regiones de Alta Baviera y Baden.[34] Si bien la variabilidad regional puede ser menos llamativa en Inglaterra y Gales, existe variación local para determinados delitos y factores de condena.[35] De manera similar, sea cierto o no, algunos jueces adquieren la reputación de ser particularmente severos con ciertas categorías de delitos, por ejemplo, violencia doméstica o conducción bajo los efectos del alcohol. Esto puede dar lugar a que los abogados busquen un juez determinado para una solicitud de libertad bajo fianza, un juicio o una vista para dictar sentencia. Los imputados que caen en una de estas culturas anómalas de imposición de penas pueden no ser conscientes de ello y, en consecuencia, no recurrir su sentencia. La IA identificaría fácilmente (y con exactitud) estas prácticas aberrantes de imposición de penas (o a los jueces atípicos), lo que daría lugar a la adopción de medidas correctoras por parte de la autoridad responsable de los lineamientos de imposición de penas, el tribunal de apelación o la instancia superior. HAL supervisaría la actuación de su empleador, el tribunal.

Todas las aplicaciones descritas hasta ahora intentan mejorar la toma de decisiones judiciales, directamente al ofrecer una SC, e indirectamente al mejorar la calidad de las propuestas. La IA también tiene un gran potencial para mejorar la naturaleza de los lineamientos existentes.

Mejora de los lineamientos para la imposición de penas

Los lineamientos para la imposición de penas se originaron en Estados Unidos en la década de 1970 y se han extendido a muchas otras jurisdicciones.[36] Aunque adoptan diferentes formas, los lineamientos comparten un objetivo común: promover una imposición

34. Grundies, Volker, "Regionale Unterschiede in der gerichtlichen Sanktionspraxis in der Bundesrepublik Deutschland. Eine empirische Analyse", en Hermann, Dieter y Pöge, Andreas (eds.) *Kriminalsoziologie: Handbuch für Wissenschaft und Praxis*, Baden-Baden, Nomos, 2018, pp. 295-313.

35. Pina-Sánchez, Jose, "Defining and Measuring Consistency in Sentencing", en *Exploring Sentencing Practice in England and Wales*, Basingstoke, 2015.

36. Roberts, Julian; Harris, Lyndon, "Sentencing Guidelines Outside the United States", en *Handbook on Sentencing Policies and Practices in the 21st Century*, New York, 2019.

de penas más consistente y basada en principios. Por lo general, los lineamientos se basan en la práctica existente y contienen recomendaciones para la imposición de sentencias que los tribunales deben seguir. Normalmente, las Comisiones o Consejos de Sentencias elaboran y publican los lineamientos. Este es el caso en Estados Unidos, Inglaterra y Gales, Corea del Sur, Escocia y otras jurisdicciones. En otros lugares, la Corte Suprema elabora los lineamientos para que los tribunales inferiores las apliquen al dictar sentencia. Hasta la fecha, la IA no ha desempeñado ningún papel en la elaboración o modificación de estos sistemas. En las jurisdicciones que aplican lineamientos, la mayoría de las aplicaciones de la IA analizadas hasta ahora –mejorar de la coherencia y la investigación sobre la imposición de penas– son ejecutadas por la autoridad competente en materia de lineamientos, la Comisión o el Consejo de Sentencias. La IA contribuiría al trabajo de la Comisión, ayudando a supervisar el cumplimiento de los lineamientos y a mejorarlos.

Identificación de áreas problemáticas en los lineamientos

La IA puede contribuir a dar forma a los mismos lineamientos. Esto podría ser identificando los aspectos problemáticos de las directrices y las formas en que los tribunales aplican erróneamente las orientaciones actuales, ya sea en los estatutos, las sentencias o los lineamientos para la imposición de penas. Actualmente esta tarea es realizada por las comisiones y los consejos de imposición de penas; la IA realizaría la misma tarea con mayor eficacia. El uso judicial de formas alternativas de custodia que se da en Canadá e Inglaterra y Gales es ilustrativo.

Muchos países aplican penas de prisión en suspenso o sanciones de arresto domiciliario. En Canadá existe la Condena Condicional de Prisión (CSI por sus siglas en inglés), que es una pena de reclusión a cumplir en el domicilio. En Inglaterra y Gales, los tribunales pueden imponer una Orden de Suspensión de la Pena (SSO por sus siglas en inglés), que también es una forma de encarcelamiento. En ambas jurisdicciones, antes de imponer una SSO o una CSI, el tribunal debe estar convencido de que la pena privativa de la libertad es inevitable. Estas sanciones sólo pueden imponerse después de que el tribunal

haya tomado la decisión de encarcelar al delincuente. Esto queda claro en una de los lineamientos.[37]

A pesar de este requisito, la investigación sugiere que los tribunales de Inglaterra, Gales y Canadá no siempre siguen esta lógica. En algunos casos, los imputados reciben un CSI o un SSO en sustitución de una pena comunitaria en vez de una pena privativa de libertad.[38] Cuando esto ocurre, se ha producido una "ampliación de la red": las formas alternativas de encarcelamiento han recurrido a la comunidad en lugar de sumar una carga de casos privativos de libertad. Determinar si, y en qué medida, los casos que reciben un CSI o un SSO estaban destinados originalmente a una pena comunitaria en lugar de la privación de la libertad es un desafío complejo de investigar.[39] Sin embargo, la IA podría identificar fácilmente los CSI o SSO con características significativamente más cercanas a los perfiles que en el pasado recibieron penas comunitarias. De este modo, sería posible estimar con mucha más precisión en qué medida se produce la "ampliación de la red", con vistas a corregir esta aplicación incorrecta de los lineamientos y el uso indebido de la sanción.

Perfeccionar los elementos de los lineamientos

Fuera de Estados Unidos, los lineamientos suelen ser específicos para cada delito.[40] Cada lineamiento contiene factores relevantes para el delito específico, y estos factores se distinguen en función de su importancia. Los lineamientos en Inglaterra requieren que los tribunales procedan siguiendo una serie de etapas, aplicando diferentes factores las distintas etapas del proceso. Por ejemplo, los

37. Sentencing Council of England and Wales, *Imposition of Community and Custody Penalties. Definitive Guideline*, 2016. Disponible en: https://www.sentencingcouncil.org.uk/wp-content/uploads/Imposition-definitive-guideline-Web.pdf [fecha de consulta: 12/04/2024].

38. Webster, Cheryl; Doob, Anthony, "Missed Opportunities: A Postmortem on Canada's Experience with the Conditional Sentence", en *Law and Contemporary Problems*, 2019, pp. 163-197; Irwin-Rogers, Keir; Roberts, Julian; "Swimming Against the Tide: The Suspended Sentence Order in England and Wales, 2000-2017", en *Law and Contemporary Problems*, 2019, pp. 137-162.

39. Webster, Cheryl; Doob, Anthony, *op. cit.*

40. Roberts, Julian; Harris, Lyndon, *op. cit.*

lineamientos ingleses asignan los factores de condena más importantes a la primera etapa de los lineamientos.

En la primera etapa, estos factores determinan cuál de los tres rangos de sentencia se utiliza. La primera etapa de la secuencia tiene el impacto más importante en la sentencia impuesta finalmente. Con el fin de promover la proporcionalidad, en la primera etapa sólo se tienen en cuenta los factores relevantes para el daño y la culpabilidad. Los factores menos importantes aparecen en la segunda etapa, donde solo afectan el desplazamiento *dentro* del rango de condena establecido en la primera etapa.[41] Por lo tanto, la "ubicación de los factores" es importante: es necesario garantizar que los factores primarios (que reflejan el daño y la culpabilidad) se asignen adecuadamente en la primera etapa, mientras que los factores menos importantes aparecen en la segunda etapa. Basta con echar un vistazo a cualquier lineamiento para darse cuenta del desafío que supone la ubicación adecuada de los factores de imposición de penas. Dos factores del lineamiento sobre robos con violación de domicilio son "ocupante en domicilio (o regresó al mismo) en presencia del delincuente" y "degradación gratuita de la víctima". ¿A qué factor debería asignársele mayor peso en la primera etapa? No está nada claro.

En la actualidad, los factores son seleccionados y asignados a las etapas según los lineamientos de los miembros del Consejo, en debate en sus reuniones. La IA puede desempeñar mejor esta tarea: podría verificar empíricamente el peso que tienen los distintos factores en sentencias anteriores, y la asignación podría reflejar estos pesos. Los factores de la primera etapa podrían definirse como cualquier factor que tenga cierto peso a la hora de predecir la gravedad de la condena. Sobre la base del análisis de la IA, el Consejo podría reasignar determinados factores, por ejemplo, si un factor de la segunda etapa resulta ser responsable por un grado muy elevado de varianza en los resultados de la imposición de penas. Por supuesto, este análisis empírico presupone que los tribunales han ubicado los factores adecuadamente hasta la fecha, y con muchas de las aplicaciones de la IA que proponemos, sería necesaria una cuidadosa revisión judicial de los resultados.

41. Roberts, Julian, "The Evolution of Sentencing Guidelines: Comparing Minnesota and England and Wales", en *Crime & Justice*, 2019, pp. 187-254.

La IA también podría identificar cualquier error general en la aplicación de los lineamientos. Uno de los rompecabezas centrales de la proporcionalidad de las penas es la relación entre los dos componentes de una sanción proporcional: ¿qué es más importante, el daño o la culpabilidad? Los lineamientos de Inglaterra y otros países están estructurados para reflejar estas dos dimensiones. Los lineamientos ingleses eluden la cuestión tratando ambas dimensiones por igual. La mayoría de los lineamientos del Consejo exigen que el tribunal asigne al caso uno entre tres niveles de gravedad al delito y uno entre tres niveles de culpabilidad. (Estos pueden concebirse a grandes rasgos como niveles bajo, medio y alto, aunque en los lineamientos se utilizan términos diferentes).

Los incrementos de gravedad siguen a ambos en igual medida. Por ejemplo, el incremento de la duración recomendada de la pena desde el nivel de daño medio al alto es de 7 años (de 9 a 16 años). El paso del nivel medio de culpabilidad al nivel más alto tiene el mismo efecto: se agregan 9 años. Pero, ¿significa esta estructura que, en la práctica, los tribunales imponen penas en las que el daño y la culpabilidad se ponderan por igual, como pretendía el Consejo de Sentencias? Los tribunales tienen mucha discrecionalidad para apartarse del punto de partida de la condena y de las extensiones recomendadas, lo que resulta, empíricamente, en que uno de los dos factores podría explicar una variabilidad significativamente mayor. O, lo que es más probable, para ciertos delitos la puntuación de culpabilidad del delincuente será mucho más determinante de la extensión de la pena que el daño; para otros delitos, podría ocurrir lo contrario. En este contexto, la IA podría determinar con precisión cómo ponderan los tribunales los dos componentes de la proporcionalidad de las penas, e identificar los delitos en los que en la práctica los tribunales se apartan de la ponderación equitativa del daño y la culpabilidad según establecen los lineamientos. En la actualidad, se desconoce en qué medida las sentencias se corresponden realmente con el equilibrio prescrito.

Identificar la frase más eficaz

Los lineamientos para la imposición de sentencias suelen ofrecer recomendaciones específicas sobre la condena. En EE. UU., éstas tienen en cuenta la gravedad del delito y los antecedentes penales del

delincuente. La IA podría sugerir una recomendación que incorporara mucha más información, por ejemplo, sobre el nivel de riesgo del delincuente y la eficacia de las sanciones alternativas. HAL sería igualmente superior en la tarea de decidir cuál entre una serie de sanciones tiene más probabilidades de mitigar la reincidencia para diferentes perfiles de delincuentes. Partiendo de una base de datos que contenga un amplio abanico de datos sobre el delincuente, su historial delictivo, el delito actual y su historial de contactos previos con el sistema penal, por citar algunas variables, la IA podría recomendar la sanción más "eficaz", del mismo modo que la IA médica predice qué pacientes tienen más probabilidades de beneficiarse de qué terapias. La IA podría ir aún más lejos. Al imponer una orden comunitaria, los jueces de Inglaterra y Gales pueden optar por imponer una o todas las condiciones o requisitos de una serie de hasta 15, como un toque de queda, un requisito de residencia o un requisito de trabajo no remunerado.[42]

Al elaborar la sentencia, los jueces intentan individualizar las restricciones impuestas, con vistas a determinar la combinación más eficaz. Esta determinación es asistida, al azar, por el abogado ("mi cliente me informa que le interesaría un programa para tratar el abuso del alcohol, y está feliz de cumplir con un toque de queda a las 7pm"), y por el consejo de los servicios de libertad condicional ("este delincuente se beneficiaría de una restricción de movilidad que le impida entrar en el centro de la ciudad o tener cualquier asociación con criminales conocidos"). En última instancia, sin embargo, es probable que las condiciones impuestas reflejen la experiencia y las intuiciones de los jueces sobre qué condiciones son más útiles o eficaces. La IA sería especialmente útil para establecer las condiciones de libertad más apropiadas para los presos en libertad condicional, cuando la decisión es puramente preventiva. De nuevo, en la actualidad, las decisiones sobre restricciones de movilidad, requisitos de presentación de informes y similares son tomadas por las juntas de libertad condicional, sin el beneficio de la inteligencia artificial.

42. Wasik, Martin, *A Practical Approach to Sentencing*, Oxford, Oxford University Press, 2014, pp. 138-150.

Conclusión

En este ensayo hemos intentado aclarar el papel adecuado para la IA en la imposición de sentencias, junto con sus potenciales aplicaciones. Compartimos la opinión expresada por otros[43] de que la sentencia es una función judicial que debe permanecer en manos humanas. Es poco probable que los imputados, las víctimas y la comunidad en general consideren la sentencia de HAL como un sustituto satisfactorio de la toma de decisiones humana. No obstante, es probable que el aprendizaje automático desempeñe un papel cada vez más importante en el apoyo de la toma de decisiones judiciales. Un algoritmo de imposición de penas proporcionaría a quienes impongan sentencias una imagen mucho más completa de las prácticas actuales, mejorando así la coherencia. En concreto, podría mitigar los sesgos cognitivos humanos y la discriminación. Ayudaría a detectar disparidades locales en la imposición de penas y facilitaría la revisión en apelaciones. Además, la IA puede mejorar los lineamientos proporcionados para ayudar a los tribunales, identificando los aspectos problemáticos de su orientación. Por último, podría fomentar una comprensión más profunda de la práctica de imposición de penas, que podría redundar sobre los elementos de los lineamientos.

¿Hasta qué punto es realista establecer un programa de IA como el aquí descrito dentro de las prácticas actuales de imposición de penas? Ya hemos señalado que HAL sería muy adecuado para llevar a cabo muchas de las tareas actualmente cubiertas por los lineamientos de imposición de penas. Por lo tanto, la transición a la imposición de penas con ayuda de la IA debería ser más fácil en aquellas jurisdicciones en las que tales lineamientos ya existan. Una evolución progresiva hacia la imposición de penas con ayuda de la IA podría comenzar, por ejemplo, con investigaciones asistidas por IA al interior de la autoridad encargada de sentenciar, y los correspondientes ajustes de los lineamientos. Sin embargo, uno de los principales obstáculos para el establecimiento de la IA será la aceptación judicial.[44] En jurisdicciones

43. Donohue, Michael, "A Replacement for Justitia's Scales?: *Machine learning*'s Role in Sentencing", en *Harvard Journal of Law and Technology*, 2019, pp. 657-678.

44. Tata, Cyrus, "The Application of Judicial Intelligence and 'Rules' to Systems Supporting Discretionary Judicial Decision-Making", en *Artificial Intelligence and Law*, 1998, pp. 203-230.

como Australia y Alemania, que sostienen firmemente un enfoque individualista (o "instintivo") de la imposición de penas, es probable que los jueces se sientan menos inclinados a aceptar la pertinencia de las penas impuestas en casos anteriores ("disímiles"). También es posible, sin embargo, que los jueces de estas jurisdicciones se sientan atraídos por la perspectiva de que HAL alcance un grado mucho mayor de individualización que los instrumentos tradicionales que estructuran la discrecionalidad. De hecho, HAL tiene el potencial de abordar algunos de los problemas constantes del enfoque "instintivo" de la imposición de penas, como la identificación de una pena adecuada de partida.[45]

Ampliar el papel de la IA en el apoyo al proceso de imposición de penas requiere una consideración cuidadosa para preservar la imposición de penas como lo que es: una empresa esencialmente humana. Sin embargo, si se emplea de forma responsable, la HAL tiene un potencial considerable para hacer que las sentencias sean más coherentes y estén basadas en principios.

45. Schöch, Heinz, "Möglichkeiten und Grenzen einer Typisierung der Strafzumessung bei Verkehrsdelikten mit Hilfe empririscher Methoden", en *Kriminologische Gegenwartsfragen*, Stuttgart, Enke, 1995, pp. 128-137.

Las sentencias algorítmicas frente al principio de culpabilidad y el derecho a ser oído en el proceso

Linus Ensel*

Reflexiones sobre su compatibilidad desde el derecho alemán. Introducción

Aunque los marcos de las penas prescritas por el Código Penal Federal alemán *(Strafgesetzbuch)* rigen por igual en todo el país, se observa una variación interregional significativa en las prácticas de imposición de penas, particularmente en lo que respecta a la severidad de las penas impuestas.[1] Un análisis de los datos de los años 2004, 2007 y 2010 revela que el 17,5% de los distritos judiciales se desvía al menos un 10% por debajo de la media federal en cuanto a la duración de las penas impuestas.[2] Cuando se seleccionan aleatoriamente dos distritos judiciales, la desviación media en la severidad de las penas para el mismo delito y la misma biografía jurídica del delincuente asciende al 15%.[3] Además, surgen disparidades en la severidad de la pena impuesta en función del cargo y la experiencia

* Investigador doctoral en el Instituto Max Planck for the Study of Crime, Security and Law de Friburgo (Alemania). El siguiente texto se basa en una presentación del 30 de marzo de 2023 en la Universidad de Buenos Aires en el marco del Coloquio Internacional de la AIDP sobre IA y Justicia Criminal. Esta contribución forma parte de una investigación en curso. El autor agradece a Dr. Ivó Coca Vila sus valiosos comentarios y correcciones.

1. Grundies, Volker, "Regionale Unterschiede in der gerichtlichen Sanktionspraxis in der Bundesrepublik Deutschland. Eine empirische Analyse", en Hermann, Dieter y Pöge, Andreas (eds.), *Kriminalsoziologie - Handbuch für Wissenschaft und Praxis,* Baden-Baden, Nomos, 2018, p. 301.

2. Grundies, Volker, *op. cit.*, pp. 297-301.

3. Ídem, p. 301.

profesional del juez.[4] Se observan fenómenos similares también en otros países.[5] La falta de estructura y uniformidad en la imposición de penas suele estar acompañada por un déficit argumental en las decisiones judiciales. La profunda transformación de la sociedad en los últimos años debido al progreso tecnológico habrá de conducir sin duda a nuevos abordajes para racionalizar y unificar el proceso de imposición de penas. En particular, se destacan dos abordajes concebibles que me gustaría tratar en mi aporte. Uno de ellos se basa en el *machine learning* (aprendizaje automático), el otro utiliza un sistema basado en reglas sin aprendizaje. En este trabajo, se examinará la compatibilidad de estos conceptos con dos aspectos centrales de la doctrina penal alemana, que también son cruciales para muchas otras jurisdicciones: el principio de culpabilidad y el derecho a ser escuchado ante un tribunal. Antes de profundizar en esta investigación, se considerarán los abordajes más tradicionales.

Abordajes convencionales

Se pueden considerar, y se han considerado, varios abordajes convencionales para reducir la disparidad de las penas. Una estrategia consiste en reducir los marcos penales. Muchos delitos del Código Penal alemán prevén actualmente un amplio marco penal, por ejemplo, de 1 a 15 años de prisión, lo que da lugar a un enorme margen a la discrecionalidad judicial. En consecuencia, una opción sería reducir

4. Streng, Frank, *Strafzumessung und relative Gerechtigkeit: Eine Untersuchung zu rechtlichen, psychologischen und soziologischen Aspekten ungleicher Strafzumessung* (R v Decker 1984), p. 118 ss., 132 s.

5. Sobre la disparidad de las sentencias en EE.UU.; Anderson, Amy L. y Spohn, Cassia, "Lawlessness in the Federal Sentencing Process: A Test for Uniformity and Consistency in Sentence Outcomes", en *Justice Quarterly*, 2010, p. 362, 383 y ss.; para un estudio en España véase Sobral, Jorge y Prieto Ederra, Ángel, *Psicología y ley: Un examen de las decisiones judiciales*, Madrid, Eudema, 1994; para Polonia véase Mamak, Kamil y otros, "A failed attempt to radically reduce inter-court sentencing disparities by legislation: Empirical evidence from Poland", en *European Journal of Criminology*, 2022, pp. 1165-1179; para la República Checa, véase Drápal, Jakub, "Sentencing disparities in the Czech Republic: Empirical evidence from post-communist Europe", en *European Journal of Criminology*, 2020, pp. 151- 165.

estos marcos, limitando así esta discrecionalidad judicial.[6] Sin embargo, los críticos argumentan que un abordaje de este tipo puede obstaculizar la capacidad de los jueces para responder adecuadamente a los casos extremos que caen en cualquiera de las puntas del espectro.[7]

Otra posible vía consiste en promulgar legislación adicional para especificar el proceso de imposición de penas. No obstante, es importante reconocer que el legislador no puede prever ni tener en cuenta todas las intrincaciones y complejidades inherentes a cada caso posible.

Por lo tanto, ha habido intentos de delegar una parte de la diferenciación del proceso de imposición de penas a un consejo o una comisión responsable del desarrollo de directrices para la determinación de la pena.[8] Su introducción se debatió entre los académicos alemanes en 2018, pero parece que las preocupaciones por la posible limitación de la discrecionalidad judicial pesaron más que la resolución colectiva para abordar la cuestión de la variabilidad de las penas.[9]

Dos abordajes modernos

¿Qué posibilidades ofrece la tecnología moderna para abordar la cuestión de la disparidad de las penas? Al considerar la imposición de penas basada en algoritmos, pueden identificarse dos abordajes distintos, cada uno con diferencias fundamentales. Aunque las dos técnicas se presentan aquí de forma simplificada, sólo es crucial comprender sus conceptos subyacentes.

El primer abordaje que aquí se presenta consiste en incorporar la inteligencia artificial, específicamente en forma de *machine learning*, al

6. Streng, Franz, *op. cit.*, 293; Sarstedt, Werner, "Referat zum 41. Deutschen Juristentag", en Ständige Deputation des deutschen Juristentages (ed), *Verhandlungen des 41. Deutschen Juristentages*, 1955, D 52.

7. Seebald, Rudolf, "Ausgeglichene Strafzumessung durch tatrichterliche Selbstkontrolle", GA 1974, p. 193-194; Bruns, Hans-Jürgen, *Das Recht der Strafzumessung - Eine systematische Darstellung für die Praxis*, 2ª ed., 1985, p. 46.

8. Como en Inglaterra y Gales y EE.UU. donde hoy en día están en vigor los directrices para la determinación de la pena.

9. *Beschlüsse des 72. Deutschen Juristentages Leipzig 2018*, Abteilung Strafrecht - A. II. 3.-7.; Kaspar, Johannes, "Sentencing Guidelines versus freies tatrichterliches Ermessen - Brauchen wir ein neues Strafzumessungsrecht?", *Gutachten C zum 72*, en *Deutschen Juristentag*, Munich, CH Beck, 2018, C 83.

sistema de imposición de penas. El *machine learning* se refiere a los procesos que permiten a los *softwares* adquirir conocimientos de forma autónoma y utilizarlos para resolver problemas específicos.[10] En este contexto, el sistema dispondría de un amplio conjunto de datos compuesto por sentencias anteriores. Analizando los hechos de cada caso y las sentencias correspondientes, el sistema "aprende" la correlación entre circunstancias específicas (*input data* [datos de entrada]) y los patrones de toma de decisiones de los jueces humanos (*output data* [datos de salida]).[11] Por consiguiente, cuando se le presente un nuevo caso, el sistema emulará el resultado de las decisiones humanas aplicando las variables aprendidas al caso actual.[12] Es importante señalar que cada decisión tomada por el sistema se basaría en decisiones anteriores tomadas por humanos. Por lo tanto, este abordaje implica determinar la pena con base en un juicio comparativo (determinacíon de la pena por contraste), en el que una sentencia se considera justa si cumple cierto grado de similitud con las sentencias impuestas anteriormente en casos similares.[13] Esta circunstancia será crucial más adelante.

En cuanto al segundo abordaje, se utiliza un algoritmo incapaz de aprender. En este caso, los criterios de imposición de penas pertinentes se implementan manualmente en el algoritmo, creando un sistema basado en reglas. Este abordaje es similar a los sistemas que emplean directrices para la determinación de la pena. El algoritmo asigna una puntuación y variables fijas a cada hecho relevante del caso.[14] Tal como los directrices para la determinación de la pena, este abordaje se centraría principalmente en factores objetivos como el daño causado y los antecedentes penales del delincuente, mientras que otras consideraciones, como las medidas preventivas, asumirían un papel secundario.

10. Ashley, Kevin D., *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge, Cambridge University Press 2017, p. 107.

11. Compárese Ashley, Kevin D., *op. cit.*, p. 107.

12. Un sistema de este tipo también es analizado por Ryberg, Jesper, "Sentencing Disparity and Artificial Intelligence", en *The Journal of Value Inquiry*, 2021, p. 3.

13. Freund, Georg, "Straftatbestand und Rechtsfolgebestimmung - Zur Bedeutung der gesetzlichen Regelungstechnik und der 'harmonisierten' Strafrahmen für die Strafzumessung", GA 1999, pp. 509-536; exhaustivamente Maurer, Matthias, *Komparative Strafzumessung - Ein Beitrag zur Fortentwicklung des Sanktionenrechts*, Berlín, Duncker & Humblot, 2005.

14. Nink, David, *Justiz und Algorithmen*, Berlín, Duncker & Humblot 2021, p. 409 ss.

Ambas estrategias son concebibles como sistemas totalmente automatizados, en los que el algoritmo llega a una decisión final de forma independiente, o como sistemas de apoyo que dejan la decisión última en manos del juez humano.[15] En esta última variante, pueden establecerse otras distinciones: el sistema puede limitarse a proponer una pena concreta o un marco penal concretizado, o puede limitar sustancialmente la discrecionalidad del juez, dejando un margen de maniobra significativamente menor para decidir.

El principio de culpabilidad

Ya examinados estos diferentes abordajes, es pertinente considerar los posibles obstáculos a su aplicación. Para ello, profundizaremos en primer lugar en el concepto de culpabilidad, que, aunque no figura explícitamente en la Constitución alemana, se basa en la garantía de la dignidad humana (artículo 1, apartado 1, de la Constitución alemana, *Grundgesetz, "GG"*) y en el Estado de Derecho (artículo 20, apartado 3, de la GG), según el Tribunal Constitucional Federal alemán (*Bundesverfassungsgericht - BVerfG*).[16] Implica la garantía de que no se impondrá ninguna pena sin la presencia de culpabilidad (*nulla poena sine culpa*). Además, según el principio de culpabilidad, la pena impuesta debe ser "justamente proporcional a la gravedad del delito y al grado de culpabilidad del delincuente".[17] En Alemania, el proceso de determinación de la pena adecuada dentro de los amplios márgenes estipulados en el Código Penal se denomina *Strafzumessung*, que se traduce literalmente como el acto de "medir la pena". El legislador alemán ha afirmado que la valoración de la culpabilidad no es una "constatación científica estricta", sino un "proceso de evaluación moral en el seno de la comunidad jurídica".[18] Algunos incluso lo han formulado como un "acto de

15. Ryberg, Jesper y Roberts, Julian, "Sentencing and Artificial Intelligence: Setting the Stage" en Ryberg, Jesper y Roberts, Julian (eds.), en *Sentencing and Artificial Intelligence*, Oxford, Oxford University Press, 2022, p. 5.

16. *BVerfGE 20, 323 – "nulla poena sine culpa"*, 331.

17. *BVerfGE 50, 5*, 12.

18. Parlamento Federal Alemán, "Entwurf eines Strafgesetzbuches (StGB) - Drucksache IV/650" 96; aprobando a Walter Grasnick, "Rationalität oder Irrationalität der

composición social".[19] Aunque un debate sobre si la culpabilidad puede medirse en absoluto excede el ámbito de esta contribución, es importante reconocer la centralidad de la culpabilidad en el marco jurídico alemán.[20] Desde el punto de vista constitucional alemán, es crucial destacar la interdependencia de la culpabilidad y la dignidad humana, tal y como se articulan de manera kantiana: La responsabilidad es la contracara de los seres humanos considerados como seres espirituales y morales que prosperan en libertad.[21] Por lo tanto, cuando se habla de culpabilidad, siempre está implícito el estatus inherente de los individuos como sujetos.[22] En consecuencia, el principio de culpabilidad salvaguarda contra la deshumanización de los individuos, impidiendo su reducción a objetos privados de todo derecho en los procedimientos judiciales.[23] La importancia del principio de culpabilidad se ve subrayada por el hecho de que los principios establecidos en los artículos 1 y 20 de la Constitución, de conformidad con el artículo 79, apartado 3 de la Constitución, están sujetos a la garantía de perpetuidad (*Ewigkeitsgarantie*), lo que significa que no pueden ser alterados.[24]

El principio de culpabilidad y los abordajes de *machine learning*

Ahora se trata de evaluar si el primer abordaje, el del *machine learning*, se alinea con el principio de culpabilidad. En primer lugar, se justifica un examen más detallado de la jurisprudencia alemana. El

Strafzumessung", en Institut für Konfliktforschung (ed), *Pönometrie: Rationalität oder Irrationalität der Strafzumessung*, 1977, p. 23.

19. Eduard Dreher, *Strafgesetzbuch mit Nebengesetzen und Verordnungen*, 37ª ed., Múnich, CH Beck, 1977, § 46 nº 12.

20. Véase, por ejemplo, el § 46, apartado 1, frase 1 del Código Penal alemán (StGB): "La culpabilidad del delincuente constituye la base sobre la que se fija la pena".

21. Kant, Immanuel, *Grundlegung zur Metaphysik der Sitten*, 2ª ed., 1786, p. 79; *BVerfGE 45, 187 – "Lebenslange Freiheitsstrafe"*, 227; Herdegen, "GG Art. 1 Abs. 1", en Dürig, Herzog y Scholz (eds.), *Grundgesetz-Kommentar*, 98ª ed, 2022.

22. Enders, Christoph, *Die Menschenwürde in der Verfassungsordnung*, 1ª ed., Heidelberg, Mohr Siebeck, 1997, p. 502.

23. *BVerfGE 27, 1 – "Mikrozensus"*, 6.

24. *BVerfGE 123, 267 – "Lissabon"*, 413.

Tribunal Federal de Justicia de Alemania (*Bundesgerichtshof - BGH*) enfatiza la necesidad de "medir la culpabilidad" *caso por caso*: "El juez debe determinar la pena apropiada en cada caso individual considerando cuidadosamente todas las circunstancias relevantes específicas de ese caso".[25] Este requisito, que limita a los jueces a considerar únicamente los hechos del caso en cuestión a la hora de evaluar la culpabilidad, contradice el concepto de determinacíon de la pena por comparación implícito en un abordaje de *machine learning*. Sin embargo, este requisito previo establecido por el BGH no puede pasar desapercibido en este contexto. No sólo se opone a un abordaje comparativo, sino que también contradice la evidencia empírica. De hecho, los jueces se basan en comparaciones con otros casos a la hora de dictar sentencia, y aproximadamente el 97% de los jueces alemanes admiten hacerlo.[26] Una encuesta entre jueces indicó que están sujetos a una significante presión de conformidad a la hora de tomar la decisión "correcta", donde esa "corrección" se refiere principalmente a adherirse al principio de igualdad y decidir dentro de los límites de las prácticas habituales.[27] Esta práctica es sensata por varias razones. En primer lugar, la aceptación social de las sentencias disminuye cuando casos similares reciben sentencias significativamente diferentes por preferencias personales de los jueces.[28] En segundo lugar, los jueces pueden temer que sus decisiones sean anuladas por tribunales superiores, ya que el propio BGH ha dictaminado que las sentencias no deben "exceder el nivel habitual para casos comparables".[29] Aunque la comparación del presente caso debe hacerse con casos "comparables", las sentencias impuestas por otros tribunales no deben tenerse en cuenta, según el BGH.[30] La intención precisa del BGH detrás de estas declaraciones aparentemente contradictorias no está del todo clara. Sin embargo, dictar sentencia

25. *BGH NJW 2011, 2597*, 2598.

26. Streng, Franz, *op. cit.*, pp. 293, 384.

27. Albrecht, Hans-Jörg, "Gleichmäßigkeit und Ungleichmäßigkeit in der Strafzumessung", en *Deutsche Forschungen zur Kriminalität und Kriminalitätskontrolle*, 1983, p. 1323.

28. Hoven, Elisa, "Die öffentliche Wahrnehmung von Strafzumessungsentscheidungen - Anlass für Reformen?", en *KriPoZ*, 17/08/2018, pp. 276-289.

29. *BGH StV 1986, 57*; *BGH, 20/10/2021 - 1 StR 136/21* no 12; *BGH NStZ 2021*, 285.

30. *BGH, 20/10/2021 - 1 StR 136/21* (n 30) no 12.

sin ningún tipo de comparación es –por las razones empíricas expuestas anteriormente– tan poco práctico como indeseable.

Realizar una comparación de casos de manera formalizada y limitada a factores específicos conlleva el riesgo de pasar por alto las características únicas de los casos. Dado que esas características únicas pueden ser factores decisivos para una sentencia proporcional, el tribunal debe tener en cuenta esas peculiaridades incluso después de haber comparado los casos. Un abordaje inicial sería aplicar la denominada "teoría del margen de maniobra" *(Spielraumtheorie)* defendida por el BGH,[31] en cuanto a preservar un margen de discrecionalidad para el tribunal de primera instancia. Teniendo en cuenta estas limitaciones, un abordaje de *machine learning* comparativo sería compatible con el principio de culpabilidad.

El principio de culpabilidad y los algoritmos incapaces de aprender

Al considerar la compatibilidad de los algoritmos incapaces de aprender con el principio de culpabilidad, el foco no se centra en la comparación, como mencionamos anteriormente, ya que un abordaje tal no implicaría un elemento comparativo. En cambio, se plantea la cuestión de si la condena esquemática en sí misma se ajusta a la noción de culpabilidad. Aquí es donde cobra relevancia la relación de la culpabilidad con la dignidad humana. Como ya se ha señalado, la doctrina alemana vincula estrechamente la dignidad humana a la prohibición de degradar a una persona a un objeto privado de todo derecho en los procedimientos judiciales. En el contexto de la tecnología moderna, algunos derivan de ello la noción de un "derecho a no ser sometido a decisiones automatizadas incontroladas de cierto alcance".[32] Algunas de las preocupaciones alemanas respecto a la "completa datificación y algoritmización" del ser humano podrían derivarse de las experiencias en el estado de vigilancia del régimen nazi.[33] Sin embargo, esta prohibición de las decisiones auto-

31. *BGHSt 7, 28, 32.*

32. Golla, Sebastian J., "In Würde vor Ampel und Algorithmus", en *DÖV* 17, 2019, pp. 673-676.

33. Ídem, p. 676.

matizadas no puede extenderse a todos los aspectos de la vida. Para ilustrar esto, se puede echar un vistazo a una decisión judicial de 1962 que examinó si la luz roja de un semáforo que indicaba la orden de detenerse violaba la dignidad humana de los afectados porque la luz roja se basaba en una decisión automatizada.[34] Está ampliamente aceptado y resulta muy convincente que los semáforos en rojo no alcanzan el umbral de la supuesta violación. Sin embargo, si consideramos la imposición de una pena, "la injerencia más grave del Estado de derecho", resulta evidente que en este contexto se ha superado el umbral. Sin embargo, si yuxtaponemos la imposición esquemática de penas y la justicia de casos individuales, queda claro que no son necesariamente contradictorias. De hecho, el código penal ya exhibe un abordaje algorítmico en forma de disposiciones del tipo "si p, entonces q": Para un delito concreto "x", se establece la correspondiente escala "y" de penas. La verdadera cuestión que se plantea, así, es la compatibilidad de un *sistema esquemático totalmente automatizado de imposición de penas* con el principio de culpabilidad. Por lo tanto, no se trata de si podríamos tener algoritmos de imposición de penas o no –dado que ya existen en forma de disposiciones penales (o incluso directrices para la determinación de la pena en algunos países)–, sino más bien de cuánto margen de discrecionalidad judicial debería preservarse para los jueces humanos una vez aplicado el algoritmo. Si la respuesta es ninguna, nos referimos a un abordaje totalmente automatizado. Entonces, ¿cuál es el problema de los sistemas esquemáticos de imposición de penas totalmente automatizados? Una vez más, la respuesta se reduce a la proporcionalidad de la sentencia. Un algoritmo de imposición de penas creado (íntegramente) por seres humanos sólo puede incorporar un número finito de consideraciones de imposición de penas. En la práctica, los algoritmos de imposición de penas existentes tienden a centrarse en un conjunto limitado de factores, como el daño causado y los antecedentes penales del delincuente. Pero como no hay dos casos exactamente iguales, el principio de culpabilidad requiere cierto margen de ajuste con respecto a las peculiaridades del caso después de considerar estos factores principales. Así, un algoritmo incapaz de aprender como abordaje totalmente automatizado no sería compatible con el principio de culpabilidad.

34. *OLG Hamburg, 18/04/1962 (2 Ss. 7/62)* 196.

Derecho a ser escuchado ante un tribunal

Profundicemos ahora en el segundo asunto, a saber, el derecho a ser escuchado ante un tribunal. Este derecho está consagrado en el art. 103, apartado 1, GG y también encuentra sus fundamentos en el principio del Estado de Derecho (art. 20, apartado 3 GG) y la garantía de la dignidad humana (art. 1, apartado 1 GG).[35] Además, el derecho a ser escuchado es equivalente a un derecho fundamental (art. 93, apartado 1, no. 4a GG). El derecho a ser escuchado abarca el derecho del acusado a expresarse ante el tribunal y el derecho a que sus declaraciones sean debidamente consideradas.[36] Sin embargo, estos dos aspectos tendrían poco valor si el acusado no pudiera familiarizarse con todos los hechos relevantes antes del proceso de toma de decisiones. Por lo tanto, el derecho a ser escuchado ante un tribunal también implica un requisito de transparencia.[37] Cabe señalar que la Constitución no especifica explícitamente la forma en que debe garantizarse el derecho a ser escuchado. En consecuencia, este derecho debe salvaguardarse tanto en los procedimientos automatizados como en los digitalizados, al igual que en los procedimientos "analógicos" tradicionales.[38]

Derecho a ser escuchado y planteamientos de *machine learning*

Ahora, el derecho a ser escuchado será evaluado en el contexto de los abordajes de *machine learning* para dictar sentencia. Es importante reconocer que nuestro principal desafío al respecto gira en torno al concepto de transparencia. Alcanzar un alto nivel de precisión en la predicción de cómo decidiría un juez humano requiere una gran cantidad de sentencias humanas para el *machine learning*. Sin embargo, un volumen

35. *BVerfGE 55, 1 – "Flughafen München II"*, 5 s.; *BVerfGE- 2 BvR 314/86* 224; Günter Dürig, "Der Grundrechtssatz von der Menschenwürde: Entwurf eines praktikablen Wertsystems der Grundrechte aus Art. 1 Abs. I in Verbindung mit Art. 19 Abs. II des Grundgesetzes", 81 *AöR* 1956, 117, 128 s.

36. *BVerfGE 7, 275 - 1 BvR 56/57 7* 278.

37. *BVerfGE 84, 188 - 1383/90)* 190; *BVerfGE 89, 28 - 1 BvR 878/90)* 35.

38. Nink, David, *op. cit.*, p. 304.

significativo de datos interconectados por variables numerosas conduce a la creación de estructuras complejas que pueden resultar difíciles, sino imposibles, de comprender plenamente para los seres humanos.[39]

Como procesado, es esencial comprender los fundamentos de una decisión judicial para poder recurrirla eficazmente. Si el propio juez no entiende cómo el modelo de *machine learning* llega a sus resultados, resulta inviable explicar el proceso subyacente a los procesados o a sus abogados defensores.

Sin embargo, también podemos cuestionarnos el grado de transparencia necesario. ¿Deberíamos aspirar a una transparencia mayor que la que ofrece el *statu quo*? Los propios jueces humanos son *black boxes*.[40] En numerosas sentencias de tribunales locales alemanes, las consideraciones sobre la condena se reducen a unas pocas frases. Un ejemplo ilustrativo, que representa numerosas sentencias, es una sentencia del tribunal de primera instancia de Düsseldorf. El condenado en este caso, cometió un hurto en circunstancias especialmente graves al apropiarse un sobre que contenía aproximadamente 50.000 euros de una habitación de hotel, mientras estaba empleado como técnico de mantenimiento. La totalidad de las consideraciones de la sentencia son las siguientes:

> La base para determinar la pena es el marco para la sentencia descrito en el § 243, apartado 1 del Código Penal, que prescribe una pena de prisión que oscila entre tres meses y un máximo de diez años por el hurto en circunstancias especialmente graves. En el caso que nos ocupa, la confesión del procesado debe considerarse un atenuante. La magnitud de los daños (50.000,00 euros) se consideró como agravante. Sopesando estas consideraciones, el tribunal impuso una pena privativa de libertad de seis meses. Esta pena es adecuada a la gravedad del delito y al nivel de culpabilidad.[41]

39. Deeks, Ashley, "The Judicial Demand for Explainable Artificial Intelligence" en *Columbia Law Review* 2019, vol. 119, p. 1829.

40. Hörnle, Tatjana "Vorüberlegungen zu Decision-Support-Systemen aus der Sicht des Strafzumessungsrechts", en Schünemann, Bernd, Tinnefeld, Marie-Theres y Wittmann, Roland (eds.), *Gerechtigkeitswissenschaft - Kolloquium aus Anlass des 70. Geburtstages von Lothar Philipps, 2005*, p. 408; Berkemann, Jörg, "Die richterliche Entscheidung in psychologischer Sicht" JZ 1971, pp. 537-538.

41. *AG Düsseldorf, 07/08/2015 - 127 Ds 51 Js 1718/14 - 69/15*, traducción del autor.

No se ha revelado el proceso que llevó al tribunal a la condena específica de seis meses a partir del rango de penas de tres meses a diez años. La falta de transparencia en este punto hace que no quede claro, especialmente para el condenado, cómo el juez ejerció la discrecionalidad judicial. En consecuencia, el proceso de toma de decisiones del juez sigue siendo opaco, parecido a una *"black box"*. Por lo tanto, cabe concluir que los problemas de transparencia no son exclusivos de los abordajes de *machine learning*. Sin embargo, sustituir una *black box* por otra no puede ser el objetivo manifiesto. Por el contrario, la falta de transparencia y la deficiencia en el razonamiento deben abordarse y rectificarse en lugar de perpetuarse. Sin duda, el concepto de Inteligencia Artificial Explicable (XAI) debe ser reconocido en esta coyuntura. La XAI es un subcampo de la inteligencia artificial, centrado específicamente en el desarrollo de modelos y algoritmos de *machine learning* que sean fácilmente comprensibles e interpretables por los seres humanos.[42] El objetivo primordial de la XAI es crear sistemas de IA que puedan ofrecer explicaciones transparentes para sus procesos de toma de decisiones.[43] El nivel de transparencia de los modelos de *machine learning* depende en gran medida de las técnicas específicas empleadas. Por ejemplo, la regresión lineal[44] ofrece un potencial considerable de explicabilidad, incluso para personas sin grandes conocimientos matemáticos.[45] Los modelos de *machine learning* más complejos, como las redes neuronales, ofrecen un grado significativamente menor de explicabilidad y son objeto de investigación en la

42. Gilpin, Leilani y otros, "Explaining Explanations: An Overview of Interpretability of *Machine learning*" en Bonchi, Francesco y otros (eds.), *IEEE 5th International Conference on Data Science and Advanced Analytics*, 2018, 80; Wachter, Sandra; Mittelstadt, Brent y Russell, Chris, "Counterfactual explanations without opening the black box: automated decisions and the GDPR" 31 Harvard Journal of Law and Technology, 2018, p. 850; Deeks, Ashley, *op. cit.*, 1834.

43. Gunning, David y otros, "XAI-Inteligencia artificial explicable" 4 Science Robotics 2019, eaay7120, 1.

44. Alpaydin, Ethem, *Machine Learning: The New AI*, MIT Press, 2016, 38; Mohri, Mehryar, Rostamizadeh, Afshin y Talwalkar, Ameet, *Foundations of Machine Learning*, 2ª edn, MIT Press, 2018, p. 275; véase también Nink, David, *op. cit.*, p. 395.

45. Borges, Georg y otros, *Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren - Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V. im Auftrag des Sachverständigenrats für Verbraucherfragen*, Sachverständigenrat für Verbraucherfragen, 2018, p. 31.

actualidad.[46] Además, hay que tener en cuenta que siempre existe una relación tensa entre la explicabilidad y la precisión del sistema.[47] Para mejorar la explicabilidad, hay que aceptar un compromiso en términos de precisión.[48] En general, cuanto mayor es el número de variables y parámetros de un sistema, más difícil resulta hacer que el modelo sea explicable.[49] Entrar en los detalles de la XAI excedería el alcance de este aporte. En esta coyuntura, conviene afirmar que la inteligencia artificial explicable (XAI) aún no ha avanzado lo suficiente como para cumplir los requisitos necesarios para su uso en un algoritmo de determinación de penas sofisticado.[50] Se ha expresado preocupación por el hecho de que hasta ahora la XAI haya sido desarrollada principalmente por entidades privadas.[51] A medida que se avance en este ámbito, lo que podría ocurrir antes de lo esperado, ciertas cuestiones planteadas por Deeks se vuelven pertinentes:

> ¿A quién va dirigida la explicación y cuán sencilla o compleja debe ser? ¿Cuánto tiempo debe tardar el usuario en entender la explicación? ¿Qué estructura o forma debe adoptar la xAI: líneas de código, presentaciones visuales, programas manipulables?[52]

La participación de agentes privados también puede causar problemas de transparencia de otras maneras. En algunos modelos de *machine learning* existentes, como los utilizados en la policía predictiva *(predictive policing)*, han surgido problemas en el pasado debido a confidencialidad de los datos subyacentes, que pueden protegerse como secretos comerciales.[53] Emplear algoritmos opacos con tanta autoridad en el ámbito de la policía predictiva ya es muy cuestionable. Pero emplearlos en el contexto de la imposición de penas podría

46. Borges, Georg y otros, *op. cit.*, p. 34.

47. Gunning, David y otros, *op. cit.*, p. 2.

48. Ídem.

49. Borges, Georg y otros, *op. cit.*, p. 31.

50. Sobre el estado actual de la investigación Lima, Gabriel y otros, "The Conflict Between Explainable and Accountable Decision-Making Algorithms", *ACM Conference on Fairness, Accountability, and Transparency* 3, 2022.

51. Deeks, Ashley, *op. cit.*, p. 1830.

52. Ibídem, p. 1837, traducción del autor.

53. Este fue el caso de infame COMPAS, véase *"State Of Wisconsin V Eric L Loomis", 881 NW2d, 749* 776.

tener consecuencias desastrosas. La utilización de algoritmos opacos no sólo violaría el derecho a ser escuchado, sino que tampoco reconocería que la imposición de penas no es (únicamente) una ciencia empírica. Más bien, el resultado de un proceso de imposición de penas, al menos en lo que respecta a la retribución, no puede corroborarse empíricamente,[54] sino que depende únicamente de su aceptación por las partes implicadas y por la sociedad en su conjunto. Revelar el recorrido hacia el resultado de la sentencia es, por tanto, indispensable para lograr esta aceptación. En consecuencia, la utilización de modelos de *machine learning* vulneraría el derecho a ser escuchado mientras estos modelos sean incapaces de autoexplicarse o carezcan de la capacidad de ser explicados exhaustivamente por otros.

El derecho a ser escuchado y los algoritmos incapaces de aprender

Los problemas de transparencia mencionados anteriormente no se plantearían en el abordaje con algoritmos incapaces de aprender. Sin embargo, en este contexto surge una preocupación distinta, conocida como sesgo de automatización. El sesgo de automatización es la tendencia a confiar excesivamente en los sistemas automatizados, lo que conduce a decisiones "que no se basan en un análisis exhaustivo de toda la información disponible, sino que están fuertemente sesgadas" hacia a las indicaciones generadas automáticamente.[55] Plantea la cuestión de si el tribunal seguirá siendo receptivo a los argumentos presentados por las partes si el algoritmo ya ha sugerido una sentencia específica. Es importante reconocer que los límites entre un abordaje totalmente automatizado y un sistema de apoyo para la toma de decisiones pueden difuminarse cuando el sesgo de la automatización lleva a aceptar la decisión generada automáticamente sin cuestionamientos.[56] En ese caso, la audiencia legal estaría siendo otorgada

54. Kaspar, Johannes, "Digitalisierung als Chance für die Strafzumessung?", en *KriPoZ*, 19/01/2023, 1, 5.

55. Parasuraman, Raja y Manzey, Dietrich H. "Complacency and bias in human use of automation: an attentional integration", en *Human Factors*, 2010, pp. 381-391.

56. Nink, David, *op. cit.*, p. 307.

por la máquina. La pregunta que surge naturalmente es: ¿Pueden las máquinas conceder audiencia legal? Mientras que la mayoría de los académicos alemanes argumentaría que sólo los humanos pueden desempeñar este papel,[57] la respuesta podría ser sólo una cuestión de progreso tecnológico. Sin embargo, es evidente que un algoritmo incapaz de aprender por sí solo, que simplemente funciona como un programa esquemático para la imposición de sentencias, no puede ofrecer audiencia, tal como una simple calculadora no puede escuchar argumentos. En cuanto a los modelos de *machine learning* y otras formas de inteligencia artificial, aún no hemos alcanzado un nivel que posibilite una interacción significativa de tal calidad. Volviendo a los algoritmos incapaces de aprender, la cuestión clave es cómo abordar el problema del sesgo de automatización. Un posible abordaje es que el algoritmo proporcione un espectro de sentencias en lugar de una sentencia específica.[58] Esto le permitiría al juez ejercer su discreción a la hora de determinar la pena exacta, teniendo en cuenta factores que el algoritmo no haya contemplado. Además, mejorar la comprensión del algoritmo por parte del juez puede fomentar el reconocimiento de sus puntos fuertes y débiles.[59] También deberían realizarse pruebas empíricas para examinar los efectos específicos del algoritmo de imposición de penas en los procesos de toma de decisiones de los jueces.

Conclusión

En resumen, la compatibilidad de los algoritmos de determinación de penas con el principio de culpabilidad y el derecho a ser escuchado depende de la inclusión o exclusión de la inteligencia artificial en los algoritmos. Si se incorpora la inteligencia artificial, la compatibilidad con el principio de culpabilidad puede lograrse garantizando que el proceso de imposición de penas vaya más allá de la mera

57. Greco, Luís, "Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf", en *RW* 2020, p. 29; Valerius, Brian, "'Legal Tech' im Strafverfahren?", en *ZStW*, 2021, pp. 152-164; Enders, Peter, "Einsatz künstlicher Intelligenz bei juristischer Entscheidungsfindung", en *JA*, 2018, p. 721; Nink, David, *op. cit.*, p. 356.

58. Ibídem, p. 308.

59. Parasuraman, Raja y Manzey, Dietrich H., *op. cit.*, p. 406.

comparación y tenga en cuenta las consideraciones de cada caso. Los algoritmos incapaces de aprender pueden alinearse con el principio de culpabilidad cuando se utilizan en sistemas de apoyo para la toma de decisiones. En cuanto al derecho a ser escuchado, los abordajes de *machine learning* aún no han alcanzado un nivel de explicabilidad que los vuelva adecuados para su aplicación en la imposición de penas. En el caso de la introducción de algoritmos incapaces de aprender, deben tomarse precauciones para mitigar el sesgo de automatización.