

Gobernanza Pública: la utilización de INTELIGENCIA ARTIFICIAL en la ADMINISTRACIÓN DE JUSTICIA. Tecnología de Innovación Pública, para la transparencia?

Rosa María Ricoy Casas
Universidad de Vigo. rricoy@uvigo.es

Introducción.

La utilización de la Inteligencia Artificial y de los algoritmos son tan frecuentes en el mundo actual, que sería difícil que un individuo pudiera pasar tan sólo unos días sin interactuar de alguna manera con uno: los estamos utilizando en los motores de búsqueda como Google; para traducir de un idioma a otro; para bloquear el correo electrónico no solicitado (spam); en los asistentes de voz; cuando usamos una aplicación móvil para encontrar la mejor manera de ir a nuestro próximo destino; en el puntaje crediticio de una persona, etc. Son de gran ayuda en casos como estos, pero cuando juegan un papel importante en la toma de decisión, por ejemplo en un proceso judicial, su utilización puede mostrar serias vulneraciones a los derechos fundamentales de la persona, incluso con una importante afectación a la libertad de un individuo, como veremos.

Como ha señalado el Consejo de Derechos Humanos de las Naciones Unidas, en numerosos países, los sistemas de asistencia y protección social se basan cada vez más en datos y tecnologías digitales que se utilizan para automatizar, predecir, identificar, vigilar, detectar, singularizar y castigar. Este proceso se conoce generalmente como transformación digital, pero no debería permitirse que este término, un tanto neutral, oculte el carácter revolucionario y la motivación política de muchas de esas innovaciones. Según algunos analistas, en el futuro los organismos públicos podrían formular leyes con robots y está claro que están surgiendo nuevas formas de gobernanza que dependen de manera considerable del procesamiento de grandes cantidades de datos digitales de todas las fuentes disponibles, utilizan análisis predictivos para prever el riesgo, automatizan la adopción de decisiones y quitan a los responsables de estas su poder discrecional. La dependencia que tienen muchos Estados de las grandes empresas transnacionales tecnológicas, y el secretismo con el que utilizan este tipo de tecnología, ya introducida en el ámbito de algunas administraciones públicas de todo el mundo, empeoran la ecuación.

Puede evidenciarse una preocupación generalizada sobre su introducción en el ámbito judicial, no sólo por parte de muchos Ciudadanos, Prensa, Abogacía, sino también por parte del Poder Judicial, las Instituciones de la UE, la OCDE, y numerosas Universidades. Este *paper* se hace eco de algunos de los principales desarrollos de la IA en el ámbito de la Administración de Justicia en todo el mundo; se exponen algunos casos jurisprudenciales sobre esta temática; las principales políticas públicas, regulaciones, y también las implicaciones que su introducción en la actualidad puede tener en los derechos fundamentales de las personas.

¿Qué entendemos por Inteligencia artificial?

El término “inteligencia artificial” (IA) *se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción -con cierto grado de autonomía- (para lograr objetivos específicos)* (COM (2018) 795 final) (COM (2018) 237 final)¹. También se ha definido como *abanico de sistemas tecnológicos que intentan automatizar o en algunos casos superar recursos computacionales y lógicos o relacionales de la inteligencia humana* (Arteaga y Ortega, 2019: 15). La inteligencia artificial reposa en gran medida sobre el aprendizaje de máquinas², es decir, el desarrollo de programas que, basándose en el suministro de información en forma de ejemplos, puedan generalizar comportamientos, de forma parecida a como aprende el cerebro humano, ¿incluso mejorándolo?

Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del Internet de las cosas) (COM (2018) 237 final). Estamos utilizando la IA diariamente, por ejemplo, para traducir de un idioma a otro; para generar subtítulos en los vídeos; para bloquear el correo electrónico no solicitado (spam); en los asistentes virtuales³; en las respuestas automatizadas; en la clasificación de imágenes; para la conversión en texto a diálogo y viceversa; en los vehículos autónomos; cuando usamos una aplicación móvil para encontrar la mejor manera de ir a nuestro próximo destino; en casa (por ejemplo, un termostato inteligente puede reducir las facturas de energía hasta en un 25% al analizar los hábitos de las personas que viven en ella y ajustar la temperatura en consecuencia)⁴; en el sector sanitario (por ejemplo, los algoritmos pueden ayudar a los dermatólogos a realizar un mejor diagnóstico -detectando el 95% de los cánceres de piel, aprendiendo de grandes conjuntos de imágenes médicas⁵-); la IA puede mejorar los productos, procesos y modelos de negocio en todos los sectores económicos; puede ayudar a las empresas a identificar qué máquinas necesitan mantenimiento antes de

¹ El Grupo Independiente de expertos de alto nivel sobre IA (IA HLEG) creado por la Comisión Europea en junio de 2018, define la IA como sistemas de software y hardware, y como disciplina científica: “Los sistemas de inteligencia artificial (IA) son sistemas de software (y en algunos casos también de hardware) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores. La IA es una disciplina científica que incluye varios enfoques y técnicas, como el aprendizaje automático (del que el aprendizaje profundo y el aprendizaje por refuerzo constituyen algunos ejemplos), el razonamiento automático (que incluye la planificación, programación, representación y razonamiento de conocimientos, búsqueda y optimización) y la robótica (que incluye el control, la percepción, sensores y accionadores, así como la integración de todas las demás técnicas en sistemas ciberfísicos)” (IA HLEG, 2019). Para más información sobre este concepto: AI HLEG (2019-2) *A definition of AI: Main capabilities and disciplines. Definition developed for the purpose of the purpose of the AI HLEG's deliverables.*

² Surgen así una serie de conceptos como son el (AA) aprendizaje automático de máquinas (*machine learning*), el (AP) aprendizaje profundo (*deep learning*) y las redes neuronales. No obstante, debe aclararse que la IA también incluye otros sub-campos que no necesariamente involucran “aprendizaje” como tal.

³ Las principales plataformas incorporan asistentes virtuales desde hace tiempo y entre los más conocidos encontramos a Siri de Apple, Google Assistant, Cortana de Microsoft o Alexa y Echo de Amazon.

⁴<https://www.la Tribune.fr/entreprises-finance/la-tribune-de-l-energie-avec-erdf/cinq-objets-connectes-poureeconomiser-l-energie-545571.html>

⁵ <https://www.theguardian.com/society/2018/may/29/skin-cancer-computer-learns-to-detect-skin-cancer-moreaccurately-than-a-doctor>

averiarse; y en general, al dar sentido a grandes cantidades de datos para ofrecer soluciones eficientes, etc.

La “electricidad” que mueve la mayoría de las tecnologías hoy en día, es la inteligencia artificial. El crecimiento de la capacidad informática, la disponibilidad de datos, y los avances en los algoritmos, han convertido la IA en una de las tecnologías más estratégicas del siglo XXI⁶. Estamos en la punta del iceberg. Esta tendencia se puede medir por la cantidad de “unicornios” (empresas de nueva creación con una valoración de al menos 1.000 millones de dólares) que emergen en determinadas regiones del mundo. El progreso en IA abre la puerta a nuevas oportunidades en áreas tales como atención médica personalizada y precisa, movilidad (conducción autónoma), tecnología financiera, manufactura avanzada, aplicaciones espaciales, redes eléctricas inteligentes, economía ecológica y biológica sostenible, mejor detección e investigación de actividades delictivas (por ejemplo, lavado de dinero, fraude fiscal), medios de comunicación, etc. Es por eso que *resulta ya insoportablemente evidente que cuando decimos que lo digital es disruptivo no nos referimos a su capacidad de aportar innovación tecnológica a los procesos o a los negocios, sino a su capacidad de transformar la sociedad. La tecnología digital está siguiendo el mismo camino que otras tecnologías anteriores como el vapor o la electricidad, y tras unas primeras etapas de descubrimiento e ingeniería ahora ya se encuentra en una fase de transformación, que se evidencia en que ya no hablamos de nuevos productos o servicios, sino de cambios de hábitos en las personas e incluso en cambios en las escalas de valores* (Roca, 2018). No obstante, la misma tecnología se puede usar con fines sociales, en beneficio de la sociedad o como en China y Rusia para un total control de la ciudadanía.

Inteligencia artificial y políticas públicas en la Unión Europea.

La nueva guerra fría es tecnológica, el nuevo escenario de confrontación es internet, e incluso de los *esfuerzos en ciberseguridad tienen vocación de defensa y también -aunque no suele reconocerse- de ataque. Estados Unidos, la primera potencia militar, apunta ya a la creación de un cuerpo de defensa espacial, al tiempo que ha intensificado las operaciones de inteligencia orientadas a contener el avance de otras potencias -China y Rusia, principalmente, pero también de la Unión Europa- en ámbitos como la tecnología 5G, la inteligencia artificial, la nanotecnología o la neurociencia* (Zafra, 2019: 6). La inclusión de la empresa china Huawei en la lista negra comercial de EE.UU. en 2019 al prohibir la venta de tecnologías estadounidenses al gigante chino de las telecomunicaciones, constituye un claro ejemplo de la creación de una de muchas “cortinas de hierro” tecnológicas que veremos en los próximos años.

El propio ciberespacio se ha convertido en un factor decisivo a la hora de analizar las relaciones internacionales. Es más, *las nuevas tecnologías basadas en el dominio de los algoritmos abrirán una brecha entre las personas y entre los países cambiando los conceptos que hoy conocemos como riqueza y pobreza*

⁶ El Grupo de Estrategia de Alto Nivel sobre tecnologías industriales ha recomendado incluir la IA como una de las tecnologías de habilitación clave debido a su potencial intersectorial crucial para la industria europea. <https://publications.europa.eu/en/publication-detail/-/publication/28e1c485-476a-11e8-be1d01aa75ed71a1/language-en>

(Olier, 2019: 60)⁷. Por poner un ejemplo, la política china denominada *Hecho en China 2025* -que pretende garantizar el dominio de China en las capacidades cibernéticas, la inteligencia artificial (IA), la aeronáutica y otros sectores de tecnología de punta- es vista no solo como un desafío económico, sino como una amenaza geopolítica. Para agravar aún más el problema, se encuentra el hecho de que China no es una democracia. El Gobierno está utilizando tecnología punta también para reprimir la disidencia y supervisar a la población (no sólo nacional).

Es, por tanto, una época de desafíos, pero también una época de oportunidades, y muchos Estados van todavía a la zaga de la innovación privada. Y es que, *el primero que ocupa una posición de ventaja tecnológica disruptiva en ella se “quedará con todo” (lo que se conoce como una estructura de winner takes all) durante un tiempo en el que puede desalojar a la competencia del mercado global y nacional (Barwise, 2018)*. Tal vez por ello, la Comisión Europea en el comunicado de prensa de 7 de diciembre de 2018 “para impulsar la inteligencia artificial fabricada en Europa”, pide que se concreten presupuestos, y estima que la UE debería gastar 20.000 millones de euros anuales (público y privado) para la I+D en IA para a próxima década. La propia Comisión, complementando inversiones nacionales, ha propuesto invertir 7.000 millones de euros en esos años (un 70% más que en el periodo 2014-2017) derivados de los programas Horizonte Europa y Programa Digital Europeo en IA.

Entre las actuaciones conjuntas para lograr el objetivo de que estas inversiones se materialicen, cabe destacar las siguientes: a) elaborar estrategias nacionales (como muy tarde a mediados de 2019 –en las que se resumirán los niveles de inversión y las medidas de aplicación-); b) una nueva asociación público-privada europea para la investigación e innovación en IA (como ejemplo, la Empresa Común ECSEL, modelo tripartito de coinversión de la Comisión, los Estados miembros y la industria para respaldar la investigación y la innovación, incluidas demostraciones a gran escala y proyectos piloto en los sectores de microelectrónica, integración de sistemas pequeños y software incorporado con un enfoque especial en proyectos de integración); c) un fondo de expansión a través del cual la Comisión preste su apoyo a las empresas de IA; d) desarrollo y conexión de centros de vanguardia europea en IA, dotándose hasta 66 millones de euros para centros de robótica.

Europa se encuentra actualmente atrasada en inversiones privadas en IA⁸, no obstante, tiene investigadores y empresas emergentes (*startups*) de clase

⁷ No hay que olvidar sucesos como el acaecido el 6 de mayo de 2010 cuando el Dow Jones cayó 1.000 puntos en tan solo unos minutos y se recuperó de nuevo unos veinte minutos después. Las razones de lo que entonces se llamó el “flash crash” se debió a ciertos algoritmos utilizados en ordenadores que operan automáticamente sin asistencia humana y compran o venden posiciones de multitud de valores en Bolsa de acuerdo a unas estrategias predefinidas. Detrás están los “quants”, unos analistas financieros que desarrollan programas informáticos con técnicas de la física o de las matemáticas, que son capaces de crear algoritmos que pueden atacar empresas cotizadas para conseguir, con la manipulación de su valor en bolsa, ingentes cantidades de dólares (Olier, 2019: 65 y 66).

⁸ Europa totalizó alrededor de 2 400-3 200 millones de euros en 2016, en comparación con 6 500-9 700 millones de euros en Asia y 12 100-18 600 millones de euros en América del Norte (McKinsey, 2017). Los Estados miembros y la Comisión también reforzarán la cooperación con el sector privado. Se espera que el sector privado se comprometa con inversiones altas y específicas en IA. Esta cooperación se basará en las asociaciones existentes en robótica y macrodatos, que representan 4.400 millones de euros en inversiones, de los cuales la mayoría (3 200 millones de euros) provienen de la industria. Las partes interesadas ya han confirmado su apoyo para establecer una asociación de IA. Las asociaciones público-privadas en robótica («SPARC») y macrodatos («Big Data Value») representan 1 200 millones de euros en inversión pública, además de los 3 200 millones de euros en inversiones privadas para 2014-2020, arrojando un total de 4 400 millones de euros. Asimismo, el sector privado podrá verse favorecido con la pretensión de la UE de fortalecer el

mundial en IA, es líder en robótica y software/plataformas de empresa a empresa. Sus fuertes sectores de transporte, atención médica y manufactura deberían estar a la vanguardia de la IA. La inteligencia artificial ocupa un lugar destacado en la agenda del Consejo de la UE desde la cumbre digital organizada por la Presidencia de Estonia en septiembre de 2017. Los principales objetivos son maximizar el impacto de las inversiones a nivel nacional y de la UE, fomentar las sinergias y la cooperación en toda la UE, incluida la ética, fomentar el intercambio de mejores prácticas y definir colectivamente el camino a seguir. Ello se pretende realizar en áreas de interés público tales como atención médica, transporte, seguridad, educación y energía, así como en otras áreas como manufactura y servicios financieros. Pero para ellos, todas las acciones concretas y complementarias a nivel de la UE, nacional y regional deben cumplir con las normas de la UE sobre el derecho de la competencia y las ayudas estatales.

En relación al ámbito público, puede destacarse que anima a las administraciones públicas de la UE a contribuir al desarrollo de soluciones y servicios habilitados para la inteligencia artificial y ciberseguridad, y a ponerlos a disposición de los ciudadanos y las empresas con el fin de mejorar la eficiencia y la eficacia de los servicios prestados, entre otros, en la contratación pública y a través de ella. Destaca la importancia de garantizar el pleno respeto de los derechos de los ciudadanos europeos mediante la aplicación de directrices éticas para el desarrollo y el uso de la inteligencia artificial dentro de la Unión Europea y a nivel mundial, y en general, subraya que toda la legislación de la UE debe ser adecuada para su finalidad y fomentar la aplicación y el desarrollo transfronterizos de tecnologías basadas en la inteligencia artificial.

La Comisión propone que se desarrollen varios “lugares de prueba de referencia a gran escala”, abiertos a todos los actores en toda Europa, utilizando hasta 1 500 millones de euros de la rama de IA del programa Europa Digital, basándose en la experiencia de los mencionados centros de excelencia existentes en los Estados miembros. El programa facilitará aún más la difusión de la capacidad de IA en cada Estado miembro y se vinculará a una plataforma de IA a petición. Entre los ejemplos de instalaciones de prueba que están implementando los Estados miembros se incluyen: Las pruebas transfronterizas de conducción conectada y autónoma, y el “Sistema Cooperativo de Transporte Inteligente (C-ITS)”. Para ello, las autoridades / operadores de carreteras están trabajando estrechamente en la plataforma C-ROADS.

Asimismo debe destacarse la experimentación a escala real de hospitales inteligentes. *Un hospital inteligente es un hospital que se basa en procesos optimizados y automatizados basados en un entorno de TIC de activos interconectados, particularmente basados en Internet de las cosas (IoT), para mejorar los procedimientos existentes de atención al paciente e introducir nuevas capacidades (ENISA, 2016)*⁹. Destaco tres aplicaciones posibles que podrían

acceso a la financiación de IA en virtud del programa InvestEU a partir de 2021. Asimismo, la Comisión está avanzando en la creación del Consejo Europeo de Innovación para apoyar las tecnologías de vanguardia y las empresas emergentes más innovadoras.

⁹ La Agencia europea para la seguridad de las redes y la información (ENISA) ha elaborado una guía con una serie de directrices de buenas prácticas que permitan salvaguardar a los llamados hospitales inteligentes: ENISA (2016). *Smart Hospitals Security and Resilience for Smart Health Service and Infrastructures*.

mejorar su gobernanza: a) la mejora de la capacidad diagnóstica y quirúrgica. La utilización de robots quirúrgicos que podrían realizar microcirugías que los médicos no pueden realizar; b) los hospitales son cada vez más capaces de extraer datos de pacientes para ayudar con el diagnóstico o elegir el mejor curso de tratamiento (incluso de forma compartida entre hospitales –con datos anonimizados–), y las sofisticadas soluciones de software les permiten ajustar sus procesos administrativos. Uno de los objetivos es la capacidad de ampliar las fronteras del hospital y proporcionar atención médica remota. Los dispositivos implantables, los dispositivos portátiles y otros dispositivos móviles introducen la capacidad de realizar un monitoreo del paciente en tiempo real a través de la medición de los signos vitales clave y hacer que estas mediciones estén fácilmente disponibles para el personal y los sistemas del hospital a través de conexiones de red (controlando las dosis médicas, proporcionando advertencias y limitando la permanencia en el hospital en aquellos casos que se consideren necesarios). También la implantación de dispositivos en el cuerpo humano¹⁰. Muchos de ellos ayudan enormemente a la investigación y el desarrollo de la medicina¹¹; c) con las TIC es posible tener información a tiempo real en todas las etapas, desde la entrada hasta la salida del paciente, la optimización de la admisión, la programación y otros procesos a su alrededor. Evitaría y, en su caso, identificaría, analizaría y resolvería cuellos de botella y, por lo tanto, contribuir a un flujo eficiente de atención médica y pacientes.

Gracias a las TICs, los siguientes dos servicios electrónicos de salud transfronterizos se están introduciendo progresivamente en todos los países de la UE: a) *ePrescription* y *eDispensation*: permite a los ciudadanos de la UE obtener sus medicamentos en una farmacia ubicada en otro país de la UE (gracias a la transferencia en línea de su receta electrónica desde su país de residencia donde está afiliado, a su país de viaje); b) *Patient Summary*: proporciona información sobre aspectos importantes relacionados con la salud, como alergias, medicamentos actuales, enfermedades anteriores, cirugías, etc. Es parte de una colección más grande de datos de salud llamada Registro de salud electrónico (*Electronic Health Record*). El resumen digital del paciente está destinado a proporcionar a los médicos información esencial en su propio idioma sobre el paciente, cuando el paciente proviene de otro país de la UE y puede haber una barrera lingüística. A largo plazo, no solo la información médica básica del Resumen del paciente, sino el Registro completo de salud podría estar disponible en toda la UE.

¹⁰ Ejemplo: los chips específicamente diseñados para ejecutar operaciones de IA (chips neuromórficos) se modelan en estructuras biológicas como los cerebros. Este proyecto forma parte de la empresa conjunta Componentes y sistemas electrónicos para el liderazgo europeo (4.800 millones de euros de inversiones público-privadas para 2020).

¹¹ Ejemplo: el proyecto emblemático sobre el mapeo del cerebro humano (*The Human Brain Project*), que reúne a neurocientíficos, expertos en informática y robótica, y es uno de los dos primeros buques insignia y una de las iniciativas de investigación más grandes lanzadas en 2013 por la UE. El proyecto central financiado por la UE coordina: 121 socios de universidades, institutos de investigación y empresas en 20 países; más de 800 investigadores en Europa y países asociados a Horizonte 2020; 18 proyectos asociados que contribuyen a la hoja de ruta de investigación estratégica. The Human Brain Project está colaborando con iniciativas en curso en el mismo dominio en Europa y más allá (EE. UU., Japón y Canadá). Forma parte de la coordinación del Área Europea de Investigación del Cerebro (EBRA) y es cofundador de la Iniciativa Internacional del Cerebro (IBI). Con ello, es posible que pueda mejorarse nuestra comprensión del cerebro, permitir un mayor desarrollo de nuevos tratamientos y diagnósticos específicos para pacientes; y desencadenar nuevos enfoques para la computación futura, incluidos los sistemas neuromórficos (inspirados en el cerebro) para IA (inteligencia artificial) y robótica.

Inteligencia artificial y Administración de Justicia.

La tecnología va muy por delante en ocasiones que el derecho y la ciencia política. Además de los ejemplos señalados en relación a las políticas públicas de IA en la Unión Europea, la introducción de la Inteligencia Artificial (IA)¹² en el ámbito judicial, promete aumentar el acceso a la justicia debido a la reducción del costo de dichos procedimientos, y muchas autoridades públicas ya han identificado beneficios presupuestarios que podrían obtenerse mediante la sustitución de algunos funcionarios judiciales por sistemas automatizados. Tal vez también porque es probable que el uso de algoritmos permita un seguimiento muy detallado y sofisticado de la actividad judicial de cada tribunal, de la ejecución de decisiones, e incluso una exhaustiva evaluación de cada juez (por ejemplo el volumen de trabajo que ha desarrollado en un plazo determinado, o el análisis de supuestos sesgos en sus patrones de comportamiento). Ello sería coherente con el cada vez más observable uso de herramientas de control laboral, cuyo perfeccionamiento está en expansión, aunque muchos también cuestionan su deseabilidad.

Existen muchas dudas también en su aplicación en torno al cumplimiento de los principios del Reglamento General de Protección de Datos (RGPD); sobre la garantía y verificabilidad de la neutralidad y objetividad de las herramientas de IA; sobre la facilidad de discernir entre los datos procedentes de su aplicación y otros datos del caso; sobre el acceso a las mismas por parte de todos los operadores jurídicos implicados y la posibilidad de discutir y cuestionar sus resultados; sobre la delegación o no del poder de decisión del Juez; y sobre la afectación a los derechos fundamentales, entre otros. Por todo ello, es necesario un profundo debate para evaluar críticamente qué papel, si lo hay, deben desempeñar las herramientas de inteligencia artificial (IA) en nuestros sistemas de justicia. La IA debe reforzar, y no disminuir, las garantías del estado de derecho y la calidad de la justicia a la que tienen acceso todos los ciudadanos.

Han surgido nuevas empresas de tecnología jurídica o tecnología legal en toda Europa (aplicados a la abogacía y el ejercicio del derecho), y para ello también se ha introducido el término *legaltech*. Las *Startups* especializadas del sector desarrollan aplicaciones cada vez más sofisticadas para la digitalización de la práctica legal por medio de *Cloud computing*, *Blockchain*, IA, *legal analytics*, *machine learning*, *Legal Gig Economy*, *Digital Selling* y *Chatbots*, etc. Sus aplicaciones son innumerables, pueden destacarse las siguientes: los *marketplaces jurídicos* o plataformas de intermediación que facilitan la captación de clientes para el profesional o despacho; los servicios de asesoramiento legal automatizado para los ciudadanos (incluso para plantear reclamaciones legales a través de internet); servicios para recopilar y generar con seguridad evidencias digitales; las herramientas para que puedan gestionar sus clientes, asuntos o minutación; para la automatización y revisión documental; para el tratamiento de grandes volúmenes de documentación que conforman algunos procedimientos; para simplificar el análisis de los datos, de la legislación, de la jurisprudencia y la

¹² En cuanto a la definición de Inteligencia Artificial (de aquí en adelante IA), la evolución histórica de su desarrollo y la agenda en la UE sobre la misma, me remito a: (Ricoy-Casas, 2019).

literatura jurídica; para el análisis predictivo de casos (posibilidades de éxito y forma de plantearlo en base a casos anteriores); la generación automatizada de documentos legales como los *Smart Contracts*, soluciones e-Discovery (identificación automatizada de documentos relevantes y revisión asistida por tecnología); los softwares de gestión y tramitación de expedientes, etc.

Sin embargo, pueden destacarse también grandes dificultades en el ámbito judicial. Por ejemplo, podría existir desigualdad de armas entre los diferentes operadores jurídicos, entre las más avanzadas capacidades que los empleados públicos (jueces, fiscales, secretarios judiciales, etc) podrían tener a su disposición, frente a los recursos limitados de los abogados (es cierto que también podría ocurrir al revés, especialmente en el caso de grandes corporaciones y bufetes de abogados). En correspondencia con ello, podrían usarse datos y elementos que no han sido objeto de un debate contradictorio por todas las partes, por ejemplo entre víctima y acusación. Todos deberían poder identificar los elementos resultantes de la implementación de una herramienta de IA. Es más, debe haber una separación estricta entre los datos o los resultados del funcionamiento de un sistema de IA y otros datos del caso.

Asimismo, no suele revelarse la manera en la que funcionan los algoritmos. Este desconocimiento en casos concretos pueden situar al acusado en una situación de indefensión, al carecer de información suficiente para poder, en su caso, impugnar las predicciones realizadas. Se socavaría el principio de imparcialidad por la imposibilidad de conocer y neutralizar los sesgos de los diseñadores del sistema. La máquina que sigue los pasos de un algoritmo para lograr un resultado, obedece instrucciones, y no hay garantía de que este proceso sea imparcial y/o justo, (e incluso puede producir errores de manera fortuita o no controlada), y hacer referencia al concepto de ética desde el diseño puede no ser suficiente.

Es conocido el caso ocurrido en junio de 2015, cuando un usuario de *Google Photos* descubrió que el programa etiquetaba a sus amigos negros como gorilas. La inteligencia artificial de Google no era capaz de distinguir una tez oscura de humano de la de simios como gorilas y chimpancés. Ese sesgo racista de la máquina obligó a disculparse a Google, que se comprometió a buscar una solución al error. Dos años después, la solución está clara: para que el programa no confunda a humanos con gorilas, han sacado a los gorilas del buscador, y a los chimpancés, y a los monos (El País, 2018) (El Mundo, 2018). Asimismo, *el primer concurso de belleza en contar con un jurado no humano, (Beauty.AI 2.0)*¹³, demostró que el algoritmo había desarrollado un sesgo racista que le llevó a puntuar más bajo a las mujeres de color (Fundación Telefónica, 2017: 15).

Así, el riesgo que pueden presentar los datos que usa la IA para aprender es también un claro ejemplo de un problema que podría afectar a la imparcialidad

¹³ Se trataba de que un jurado formado por robots, observaran las fotos enviadas de personas tomadas por una aplicación estandarizada, evaluando las características principales, por ejemplo, el color de piel, las arrugas, las asimetrías, etc. Parámetros que afectan a la percepción de la belleza. Algunas de las reglas eran no maquillarse, no portar gafas, o no presentar barba: <http://beauty.ai/>

de un juicio. Los sistemas de IA no comprenden todo el contexto de nuestras sociedades complejas. Sus datos de entrada son su único contexto para ellos y si los datos proporcionados para entrenar a la IA son parciales o incluyen sesgo (incluso no intencional), entonces se puede esperar que la salida de la IA sea incompleta y sesgado también. En algunos estados de los Estados Unidos, existen sistemas de inteligencia artificial que analizan la probabilidad de que un presunto delincuente cometa delitos si el infractor es puesto en libertad bajo fianza (se utilizan para tomar la decisión de otorgar o no una fianza). Sin embargo, tales sistemas han sido acusados de discriminar a sectores particulares de la sociedad, no por motivos inherentes a la debilidad en los algoritmos, sino porque el conjunto de datos utilizado para educar al sistema contenía sesgos ocultos.

De hecho, *la empresa Northpointe*, que creó la aplicación COMPAS, *desarrolló un algoritmo para predecir la probabilidad de que un delincuente reincidiera y este empezó a juzgar que las personas de color eran más proclives a volver a cometer un crimen que los blancos* (Fundación Telefónica, 2017: 14). Con este precedente, y con las estadísticas que hay en algunos países sobre delincuencia negra, cuesta mucho trabajo pensar que no es uno de los factores utilizados para “entrenar el algoritmo”, o cuyo sesgo se evita por muchos programas de este tipo. Algunos autores han confirmado la utilización de este sesgo, señalando que, tras una investigación demostraron que los acusados blancos fueron etiquetados erróneamente como de bajo riesgo con más frecuencia que los acusados negros (Angwin, Larson, Mattu, Kirchner, 2016).

Por tanto, la utilización de este tipo de tecnología puede exacerbar disparidades injustificadas e injustas que ya son demasiado comunes en el sistema de justicia penal y en nuestras sociedades. Es más, aun no siendo un sesgo directo, algunos autores como Harcourt señalan que con el uso de estos instrumentos, el riesgo es un "sustituto de raza", al dar un gran peso a los antecedentes penales, que están altamente correlacionados con la raza. Argumenta que esta estrategia indudablemente agravará el ya intolerable desequilibrio racial en las poblaciones de las prisiones de algunos países como Estados Unidos (Harcourt, 2010). El sesgo racial ya se ha visibilizado también en otros ámbitos como el de salud.

A estos problemas, se añade el de la incertidumbre sobre su funcionamiento real. Los modelos de aprendizaje automático, al comportarse en cierta medida de manera autónoma, siguen siendo en su mayoría “cajas negras” (*black boxes*). Aunque se elaborase un manual para detallar los datos incorporados, la forma en que se ha programado, ni siquiera en este caso podría conocerse con detalle cómo gestiona la información. Ni siquiera los expertos en el campo de la IA pueden ser capaces de prever las "decisiones" tomadas por un sistema de IA, o explicar el proceso por el cual esas decisiones fueron tomadas. Por tanto, se evidencian brechas importantes, no es posible su implementación con garantías, y por ende, su aplicación en este momento haría más vulnerable la defensa de los acusados. De lo contrario, de usarse estos sistemas en la actualidad, se estarían utilizando resultados que van más allá del razonamiento humano, lo cual podría conducir a decisiones de la Corte mal justificadas y motivadas, limitando

el derecho a la defensa. Explicabilidad, transparencia, trazabilidad y equidad son algunos de los elementos que deben perfeccionarse para la implementación de la IA, especialmente en justicia.

Otras conclusiones

Es importante que las administraciones públicas en Europa sean líderes en el uso de la IA, pero es *imprescindible que las instituciones públicas adopten una estrategia proactiva, es decir, que aprovechen la revolución tecnológica no solo para renovar su capacidad técnica sino especialmente para solventar la mayor parte de sus problemas conceptuales y organizativos. La inteligencia artificial y la robótica pueden ser la gran oportunidad para implantar una renovación institucional y organizativa radical de las instituciones públicas y contribuir a su adaptación, a su relevancia y a su supervivencia en un contexto complejo de gobernanza. El objetivo es ir más allá de la simple digitalización e innovación tecnológica y aprovechar el nuevo paradigma tecnológico para ordenar los distintos modelos y culturas de la Administración (burocrático, gerencial, regulador y de gobernanza) y conseguir una mayor fortaleza institucional* (Ramió Matas, 2018: 7). No obstante, no puede olvidarse que estos Entes están obligados a actuar según lo prescrito por la ley, que necesitan motivar sus decisiones y que sus actos están sujetos a revisión judicial por los tribunales administrativos.

Todo ello debería abrir nuevas posibilidades de participación ciudadana a todos los niveles, incluido, por ejemplo, el del diseño del futuro de la Unión Europea, no desde las élites y los Estados sino también desde abajo, desde los ciudadanos, y no presentarse la digitalización de la sociedad, y la incorporación de la IA como un riesgo de exclusión social. Con respecto a los resultados que se pueden obtener con ellos, puede señalarse una serie de *métricas para identificarlos como el número de personas que accede a la plataforma, la diversidad sociodemográfica y ubicación geográfica de los participantes, el número de comentarios o participaciones ciudadanas que se logran en la plataforma y la incorporación de los comentarios ciudadanos a los proyectos de ley* (Tello, 2016).

Cada usuario del sistema, debe ser informado, en lenguaje claro y comprensible de la naturaleza vinculante o no vinculante de las soluciones propuestas por los instrumentos de IA, las opciones posibles y su derecho de asesoría legal y el recurso ante un tribunal. Debe poder identificarse los datos obtenidos por IA, y verificar su funcionamiento -debe existir la posibilidad de verificar la entrada de datos y el razonamiento de la herramienta de IA (que debe ser transparente, sin errores ni sesgos, con el pleno cumplimiento del RGPD), y sin que pueda delegarse el poder del Juez en este tipo de tecnología. Ello hará posible también la posibilidad de discutir y cuestionar los resultados de la IA. Como muestran muchos autores e informes técnicos, todavía quedan muchos interrogantes que responder antes de su aplicación.

Las soluciones que puedan encontrarse, no pueden depender en este ámbito de la relación coste-beneficio, ni en la eficiencia basada en la evidencia numérica más que de juicios individuales. Los instrumentos no proporcionan nada parecido a las predicciones precisas del riesgo de reincidencia individual. Los modelos de regresión subyacentes pueden proporcionar estimaciones razonablemente precisas de las tasas medias de reincidencia del grupo de infractores. Pero la justicia, la imposición de sanciones no puede justificarse sobre la base de estadísticas y generalizaciones sobre las tendencias del grupo.

Bibliografía

AI HLEG (Grupo independiente de expertos de alto nivel sobre inteligencia artificial creado por la Comisión Europea) (2019) *Ethics guidelines for trustworthy AI*.

AI HLEG (Grupo independiente de expertos de alto nivel sobre inteligencia artificial creado por la Comisión Europea) (2019-2) *A definition of AI: Main capabilities and disciplines. Definition developed for the purpose of the purpose of the AI HLEG's deliverables*.

Angwin, J., Larson, J., Mattu, S., Kirchner, L., «*Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*», en ProPublica, (2016), 23/03/2016, disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Arteaga, F y Ortega, A. (2019). *Hacia un ecosistema español de Inteligencia Artificial: una propuesta*, Real Instituto Elcano, Madrid, España. http://www.realinstitutoelcano.org/wps/portal/rielcano_es/contenido?WCM_GLOBAL_CONTEXT=/elcano/elcano_es/zonas_es/policy-paper-2019-hacia-ecosistema-espanol-inteligencia-artificial-una-propuesta

Barwise, P. (2018), "Nine reasons why tech markets are winner-take-all", disponible en: <https://www.london.edu/lbsr/nine-reasons-why-tech-markets-are-winner-take-all>

El Mundo (2018). *Google soluciona a las bravas su algoritmo 'racista', que confundía a personas negras con gorilas*: <https://www.elmundo.es/f5/comparte/2018/01/16/5a5dff50468aeb81638b45e7.html>

El País (2018). *Google arregla su algoritmo "racista" borrando a los gorilas*: https://elpais.com/tecnologia/2018/01/14/actualidad/1515955554_803955.html?id_externo_rsoc=TW_CM;

ENISA (2016). *Smart Hospitals Security and Resilience for Smart Health Service and Infrastructures*.

Fundación Telefónica. (2017). *Inteligencia artificial. Las máquinas que aprenden solas*. <http://fundaciontelefonica.com/>

Harcourt, B.E., (2010). Risk as a Proxy for Race, en *Criminology and Public Policy, Forthcoming, University of Chicago Law & Economics Oline Working Paper n.535, (16/09/2010), University of Chicago Public Law Working Paper n.323*, disponible en: <https://ssrn.com/abstract=1677654>

Mckinsey (2017). 10 imperativos para Europa en la era de la IA y la automatización, citado en *Comunicación de la Comisión al Parlamento europeo, al Consejo europeo, al Consejo, al Comité económico y social europeo y al Comité de las regiones. Plan coordinado sobre la inteligencia artificial*. COM (2018) 795 final.

Olier, E. (2019). Algor, el poder del control social, en *Telos* 110

Ramió Matas, C. (2018). Inteligencia artificial, robótica y modelos de Administración pública, en *Revista del CLAD Reforma y Democracia*, No. 72.

Ricoy-Casas, R.M. (2017). Las políticas sociales a través de las Smart Cities, en: Herrera-Gutiérrez, M.R. y Jaráiz, G. *Pactar el futuro: Debates para un nuevo consenso en torno al bienestar*, Disponible en: <https://rio.upo.es/xmlui/handle/10433/4336>

Ricoy-Casas, R.M. (2019). Inteligencia Artificial y políticas públicas en la UE, en: Valcárcel Fernández, P.; Fernández Acevedo, R. y Bonorino Ramírez, P.R. (dirs.): *Derecho, desarrollo y nuevas tecnologías*, Aranzadi Thomson Reuters, ISBN: 9788413087184 9788413087207

Ricoy-Casas, R.M. (2021). Utilización de medios tecnológicos y datos personales en las actividades electorales en España, en: García Amado, J.A. y Sendián Mateos, J.A. (dirs.): *Argumentación jurídica y conflictos de derechos*, Tirant lo Blanch, ISBN: 978-84-1378-420-5

Ricoy-Casas, R.M. (2021). *inteligencia artificial y administración de justicia: una política pública sub iudice*, Aranzadi Thomson Reuters, (en prensa).

Roca, Genís (2018), Sobre disrupción digital, leyes y taxis, en *La Vanguardia*, <https://www.lavanguardia.com/local/barcelona/20180730/451165907765/sobre-disrupcion-digital-leyes-y-taxis.html>

Tello, Irene (2016), “*Crowdlaw*: incorporando a los ciudadanos a la creación de leyes”, en *Este País*, México, <http://www.estepais.com/articulo.php?id=687&t=crowdlawincorporando-a-los-ciudadanos-la-creacion-leyes>.

Zafra, J.M. (2019) La nueva guerra fría es tecnológica, en *Telos*, nº110.