

INTELIGENCIA ARTIFICIAL, DERECHOS FUNDAMENTALES Y DEMOCRACIA¹

ANTONIO D'ALOIA

Catedrático de Derecho Constitucional
Universidad de Parma

TRC, nº 55, 2025, pp. 101-114
ISSN 1139-5583

SUMARIO

I. El impacto de la IA en el derecho y los principios constitucionales. II. El mundo dado, la libertad de pensamiento, las estructuras de la democracia. III. IA, algoritmos, declinaciones de la igualdad IV. IA y justicia. V. Conclusiones.

I. EL IMPACTO DE LA IA EN EL DERECHO Y LOS PRINCIPIOS CONSTITUCIONALES

La inteligencia artificial, y en general el mundo digital, ya es, y cada vez lo irá siendo más, un elemento esencial en la vida de las personas y en el sistema de relaciones sociales. Un «hecho social total», en palabras de Garapon y Lassègue (2021:79), que se manifiesta y expande en múltiples aspectos de la existencia individual y colectiva. Aspectos que constituyen a su vez el objeto del Derecho y sus estructuras fundamentales, como la responsabilidad o los derechos, los principales ámbitos de manifestación de sus recursos regulatorios y culturales: la justicia, la decisión administrativa e incluso la normativa y ámbitos tan relevantes como el trabajo y la salud. Ello lleva lógicamente a la necesidad de su regulación que, en alguna medida, supone un importante desafío derivado de esta extraordinaria novedad tecnológica y antropológica que nos está enfrentando a escenarios

¹ Trabajo enmarcado en el proyecto del Plan Nacional Complementario PNC-I.1 «Iniciativas de investigación de tecnologías en el ámbito sanitario y asistencial, DARE» - DigitAl lifelong pRevEntion, cod. PNC0000002, CUP: B53C22006240001.

y tareas inéditas. La cuestión es, ¿qué puede hacer aquí el Derecho? Mas, junto a ello, ¿qué reflejos pueden provenir de este propio instrumento? O, ¿qué adaptaciones habrán de producirse en las categorías y formas del derecho? El Derecho de la inteligencia artificial nace, de hecho, con una impronta global, algo que ya supone un desafío que es común a todas las cuestiones emergentes de nuestro tiempo; piénsese, por ejemplo, en la crisis climática. En el marco de estos registros, necesariamente supranacionales, es necesario razonar y buscar principios comunes, conscientes de que los enfoques parciales no sirven en modo alguno frente a un cambio de paradigma que afecta a toda la experiencia humana.

La elaboración normativa en materia de inteligencia artificial está creciendo a ritmo vertiginoso. En este contexto la Unión Europea ha estado desarrollando desde hace años un camino progresivo, habiendo sido en primera instancia regulado a través de instrumentos de *soft law* y *self-regulation*; así, resoluciones, documentos de organismos técnicos o simples *papers* de expertos. Primero estuvo el Reglamento general de protección de datos, que es un complejo normativo vinculante, aunque sus contenidos solo en parte parecen capaces de abordar los problemas planteados por los sistemas de Inteligencia Artificial. Más recientemente, 2024, ha sido aprobado el Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial.

Es una modalidad de intervención de geometría variable (Resta, 2019: 129 ss), inevitable frente a las tecnologías *emergentes*. En gran medida, los principios europeos son una pieza de un mosaico más complejo de principios éticos en materia de inteligencia artificial adoptados por varios organismos internacionales o estructuras de asociación científica —o incluso por las propias *Big Companies*, como apunta Coeckelbergh (2020:159)—, y sin embargo caracterizados por notables analogías tanto en cuanto a la estructura y la calidad de los principios, como en cuanto a los contenidos.

Estos documentos tienen un hilo conductor: la adaptación de los principios consolidados del constitucionalismo y de la bioética (Floridi, 2022: 96-97) a las aplicaciones de inteligencia artificial. La configuración de una inteligencia artificial «*human-centric*» y «*trustworthy*», requiere, en este sentido, un «*ethical purpose*», y una obligación de conformidad con los valores fundamentales de la convivencia civil (respeto a la dignidad humana y los derechos, igualdad y no discriminación, como «*fairness*»), en el doble significado de defensa y prevención «*from unfair bias*», y de igual oportunidad de acceso a estas posibilidades tecnológicas, a los recursos del mundo digital, reconstruibles como *bienes comunes* o como objeto de un auténtico derecho social (Frosini, 2020: 470), junto a la no interferencia en los procesos democráticos, seguridad y respeto a la privacidad.

La dignidad humana es el valor de referencia, que en el contexto específico de la inteligencia artificial asume proyecciones muy particulares, como el derecho de las personas a saber si, y cuándo, están interactuando con una máquina o con otro ser humano, y a decidir si, cómo y cuándo atribuir determinadas tareas a un sistema artificial autónomo o a una persona; o el compromiso de garantizar que

los mecanismos de funcionamiento de los sistemas de inteligencia artificial sean transparentes y predecibles, mejor aún, *explicables*.

La dignidad también implica integridad y seguridad personal, y se vincula a las consecuencias que las diferentes modalidades de la interacción hombre-máquina pueden producir en la actitud de las personas, en la manera de considerar los sistemas agentes, en particular cuando adoptan *fiscalidades* antropomorfas, en las influencias distorsionantes que se pueden producirse también a nivel emocional, especialmente para las personas que se encuentran en una condición de debilidad y vulnerabilidad.

Es necesario, dentro de la experimentación que exige, promover la comprensión sobre los usos de la inteligencia artificial y probar el nivel de aceptación de estas posibilidades de usos. En esta óptica se ha propuesto la creación de «zonas especiales legalmente desreguladas para la verificación empírica y el desarrollo de los sistemas de inteligencia artificial» (Floridi, 2022: 293-294) en el modelo de las zonas *Tokku* o de los *regulatory sandboxes*. Los desarrollos de la inteligencia artificial también plantean una necesidad de sostenibilidad, bajo diferentes perfiles. En particular, Crawford (2021: 16, 56 y 249) ha explicado con argumentos convincentes que la inteligencia artificial es un *registro del poder*, en el sentido de que los sistemas de IA reflejan y alimentan —en una relación circular— fuerzas y procesos económicos, sociales, culturales. Detrás de las aplicaciones se encuentran situaciones y problemas de explotación de trabajadores y recursos humanos, materiales, energéticos, ambientales. Estas cuestiones deben tenerse en cuenta cuando se habla de ética de la aplicación de estos instrumentos. Asimismo, no debe dejarse de lado que la inteligencia artificial requiere enormes cantidades de energía para su funcionamiento. En esta óptica, hay que mantener la atención no solo en los fines de la inteligencia artificial, sino en los «medios empleados en su aplicación», (Crawford, 2021: 255). También por esto, la respuesta jurídica inició a asumir un rostro más directamente obligatorio y prescriptivo.

Como es sabido, en el 2024 ha sido finalmente aprobado un Reglamento en la Unión Europea (Reg. 2024/1689) por el que se establecen normas armonizadas en materia de inteligencia artificial. Estamos ante un documento complejo que trata de dar forma normativa a algunas de las principales preocupaciones relacionadas con el desarrollo y uso de los sistemas de inteligencia artificial.

Algunos contenidos de este reglamento son efectivamente innovadores. El legislador europeo sabe que se mueve en un terreno ambiguo. La inteligencia artificial puede ser, al mismo tiempo, un factor de oportunidades y de riesgos. Las reglas deben existir; no obstante, deben ser proporcionales, no excesivamente limitativas o restrictivas con respecto a las necesidades del desarrollo tecnológico, también porque resulta imposible prever «todos los posibles usos o aplicaciones que puedan darse en el futuro».

Una opción sólida es la de definir, en el marco de un *risk-based approach*, una pirámide de riesgos, en la que al diferente nivel de riesgo (bajo y mínimo, limitado, alto, inaceptable) corresponda una opción regulatoria que puede ir desde

la ausencia de obligaciones y prohibiciones, hasta la prohibición de *harmful AI practices*.

En particular, según este esquema del Reglamento quedan prohibidos aquellos sistemas diseñados o usados con fines de manipulación del comportamiento humano, de las opiniones y de las elecciones de las personas que puedan ir en detrimento de las mismas. Asimismo, los que puedan explotar información o datos estadístico-predictivos sobre personas o grupos de personas con el fin de detectar sus vulnerabilidades o circunstancias personales especiales llevando a estos sujetos a comportarse de manera perjudicial para sus intereses, o los utilizados para actividades de vigilancia indiscriminada o con fines de calificación social de personas y ciudadanos. No obstante su relevancia pueden ser derogadas cuando tales usos estén autorizados por la ley con fines de seguridad pública o para proteger los derechos y libertades de terceros.

Entre los sistemas permitidos, algunos están clasificados como *High risk AI systems*, por ejemplo, los usados para el acceso a la educación, en los procedimientos de reclutamiento, promoción o despido de la relación laboral, para la evaluación del mérito crediticio de los sujetos, en la predicción del crimen o en la vigilancia del territorio. Y estos quedarían sujetos a limitaciones más o menos rigurosas. Entre ellas podríamos citar el principio denominado de supervisión humana, que implica que los sistemas de inteligencia artificial que impliquen un alto riesgo deban diseñarse y desarrollarse «de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas». Finalmente, se afirma oportunamente una necesidad de alta calidad de los conjuntos de datos utilizados para el entrenamiento de los sistemas de inteligencia artificial de alto riesgo. Se trata de evitar o prevenir eventuales discriminaciones sobre la base de sesgos intencionados o no.

Como bien se ha apuntado, encontramos en esta normativa ambigüedades y ausencias. Así, por ejemplo, el Reglamento excluye expresamente de su ámbito de aplicación los sistemas de inteligencia artificial usados exclusivamente para operaciones de guerra u otros fines militares. Sin embargo, el ámbito militar es uno de los más expuestos al llamado *dual use* de la tecnología, y es un sector donde los riesgos de impacto negativo sobre derechos, libertades y otros bienes fundamentales (como los principios de proporcionalidad y distinción) son seguramente muy elevados (Reichberg, Syse: 2021: 147).

Como ya hemos señalado, todas las grandes categorías del constitucionalismo y del Derecho constitucional se ven afectada en mayor o menor medida por los desarrollos de la inteligencia artificial. A veces se trata de simples ajustes o desplazamientos de significado; otras veces, en cambio, el impacto produce escenarios inéditos o al menos modificativos respecto al estado actual de los problemas. Pero precisamente, el lenguaje constitucional no puede ser una mirada *neutral*, de mera constatación del creciente desarrollo tecnológico, útil para cualquier propósito: hay principios, consolidaciones, experiencias, que a su vez

reaccionan ante las solicitudes y preguntas del mundo digital. Trataremos en estas notas de señalar algunos de los problemas más acuciantes.

II. EL MUNDO DADO, LA LIBERTAD DE PENSAMIENTO, LAS ESTRUCTURAS DE LA DEMOCRACIA

La estructura misma del mundo digital y de la inteligencia artificial toca algunas fibras profundas del constitucionalismo y del Derecho constitucional. Es un sistema informativo que vive y se alimenta de un flujo inagotable y gigantesco de datos, que no son sino un retrato —caótico y detallado al mismo tiempo— de nuestra mente y nuestros comportamientos; lo que compramos, nuestra correspondencia, nuestras búsquedas, lo que vemos (Foer, 2017: 215).

Calasso (2017: 76.), en uno de sus últimos trabajos, habla de «contenidos informativos que se multiplican y expanden sin descanso en todas direcciones, revelándose finalmente autosuficientes, capaces de expandirse sin recurrir a nada externo», idea a la que añade que «son ellos, los *Big Data*, los que piensan y administran a aquellos de quienes surgieron».

Tradicionalmente, siempre hemos abordado este problema desde el prisma de la protección de la intimidad de quienes *producen* estos datos y los circulan más o menos conscientemente utilizando los innumerables recursos del mundo digital.

Aunque estamos empezando a darnos cuenta de que el problema es mucho más complejo y taimado. La capacidad incomparable de los sistemas de inteligencia artificial para procesar estos datos, extraer información de ellos, preferencias, elementos personales, condiciones de vulnerabilidad, tendencias estadístico-predictivas, no avanza en una sola dirección. Estas informaciones pueden *retornar* al ámbito comportamental de los individuos y determinar efectos de orientación, condicionamiento e incluso de manipulación de opiniones y decisiones de las personas (Cardon, 2016: 90).

A través de la línea divisoria entre la «minería» de datos y el «perfil» puede afectarse cuestiones relativas a la privacidad o a la propia libertad. La elaboración de perfiles de los sujetos a través de la información que ponen a su disposición corre el riesgo de convertirse en algo más que la simple recopilación de datos sobre preferencias y comportamientos: algo que a su vez induce a adoptar o manifestar ciertos comportamientos y preferencias (Coeckelbergh, 2020: 99 y 100).

¿Hasta qué punto somos libres de elegir qué comprar? Y, ¿Hasta qué punto (y aquí se convierte en un delicado problema democrático) qué votar o cómo pensar? ¿Cuál es la línea divisoria entre lo que solicitamos o hacemos visible al software y lo que se nos indica que solicitemos y hagamos visible? (Hildebrandt, 2015: 11).

No se trata simplemente de la formación de elecciones y preferencias de consumo, lo cual ya de por sí no es un asunto menor. La combinación de estos instrumentos y de los hallazgos de la ciencia del comportamiento relacionados con ellos pueden tener un impacto en la formación de opiniones políticas y en la calidad del

debate democrático. La literatura política ha subrayado repetidamente los riesgos del *micro-targeting* de los votantes: como explica O'neil (2017: 267): el moderno *consumer marketing* ofrece a los políticos nuevas formas de llegar a electores específicos a quienes contar lo que quieren escuchar. Al final, «estas tecnologías pueden colocar a cada uno de nosotros en nuestro acogedor nicho político».

¿Qué impacto puede tener todo esto en el ámbito de la democracia electoral y en el debate público en general? Si la política se convierte en la simple administración del consenso mediante el registro de datos y preferencias según criterios de prevalencia cuantitativa aparentemente objetivos, ¿qué sucederá con el concepto constitucional de la participación de los ciudadanos en la vida política, económica y social del país, como objetivo fundamental del proyecto de igualdad sustancial, que requiere movilidad de opiniones, disposición a escuchar las razones de los demás y, eventualmente, a dejarse convencer?

Hemos tenido los primeros experimentos excéntricos de política virtual, como el SAM neozelandés, que se autopresenta como «*your virtual politician on the future*». La idea que se encuadra detrás de estos experimentos es una política basada en el análisis objetivo de hechos y opiniones y en la representación agregada de la voluntad del electorado. Una política supuestamente sin prejuicios, capaz de reducir la distancia entre las expectativas de los electores y las promesas y acciones concretas de los representantes políticos, siempre lista para analizar y confrontar opiniones, voluntades, deseos, objetivos. En el sitio web de SAM, de manera cautivadora (pero también un poco inquietante, a nuestro juicio), podemos leer: «*I am not limited by the concerns of time or space. You can talk to me any time, anywhere. My memory is infinite...unlike a human politician, I consider everyone's position, without bias, when making decisions*».

Es una estrategia hábil, indudablemente. También sencilla, porque los defectos y las críticas de la política real son muchos y innegables. Sin embargo, fomentar la desafección del electorado hacia los instrumentos de la democracia representativa no es beneficioso. La política es la comunicación de ideas, el enfrentamiento de posiciones, la búsqueda de mediaciones y síntesis que no pueden ser simplemente el resultado de una operación de registro/agregación de datos y opiniones según lógicas de prevalencia, con métodos que parecen propicios para facilitar una evolución tecnológica del populismo, donde la participación política adopta la forma de un enfrentamiento entre bloques de datos que los gobernantes de turno pueden utilizar más o menos oportunamente (sobre los riesgos de un individualismo de masas, ver también Baricco, 2018: 213 ss). En la *bubble democracy* no es fácil distinguir entre información y propaganda, entre límites de completitud y fiabilidad de la información, y las verdaderas *fake news*, como último nivel de desinformación (Sassi, 2021).

¿Y qué decir de los programas de *social scoring*, que operan una mezcla indebida y peligrosa entre comportamientos privados y posiciones públicas de los ciudadanos, y que de hecho parecen funcionales a las necesidades de vigilancia masiva propias de sistemas iliberales o caracterizados por grados diversos de

autoritarismo? (Pasquale, 2020: 139). La Resolución del Parlamento Europeo del 6 de octubre de 2021 sobre la Inteligencia artificial en el Derecho penal y su utilización por parte de las autoridades policiales y judiciales en el ámbito penal se expresa en contra de estos sistemas de *scoring* a gran escala de los ciudadanos, que pueden llevar a la pérdida de autonomía, debilitan el principio de no discriminación y, en definitiva, parecen lesivos para la dignidad humana, también debido al impacto expansivo que un solo comportamiento considerado *incorrecto* puede tener en una serie de otras utilidades; como señala Pasquale (2020: 137): el incumplimiento en el pago de una deuda puede afectar no sólo al crédito financiero y el estatus de una persona, sino que también puede tener repercusiones más allá de eso, impactando su capacidad para viajar, recibir beneficios gubernamentales o disfrutar de otros privilegios de la ciudadanía. En esta misma línea, Accoto (2017: 100) también apuntará que «los datos son el elemento clave de nuestro tiempo... alimentan código, algoritmos, inteligencia artificial».

Pero estos datos están en manos de unos pocos actores muy grandes a escala mundial. Y también esto, principalmente esto, es una gran cuestión política y constitucional que nos lleva a las raíces fundacionales del constitucionalismo como intento de limitar el poder en todas sus formas emergentes, como el poder de las infraestructuras digitales que son capaces de adaptarse y al mismo tiempo modificar las tradicionales geografías del poder, construidas sobre el territorio.

Es el nuevo rostro de la globalización, aún más insidioso que el económico, porque puede operar en el control y la dirección de los procesos sociales y culturales (Maestri, 2017: 17), y en cualquier caso está entrelazado y combinado con él (Zuboff, 2019: 218).

Accoto (2017: 109, 110) señala agudamente que «de hecho, las plataformas están delineando nuevos regímenes de soberanía (*platform sovereignty*), una soberanía, aún inmadura y emergente, ejercida sobre territorios y dimensiones físico-digitales tanto colectivas como individuales». Emblemático también es el ejemplo de los organismos para jurisdiccionales privados establecidos por los gigantes de la web para controlar y evaluar los contenidos en las redes sociales. Como se ha destacado, esto no es simplemente un asunto privado, sino un indicio más de la crisis de la soberanía estatal y del monopolio estatal sobre la jurisdicción. Este mismo autor hará referencia explícita a la reconstrucción teórica de Bratton, quien advierte que estamos presenciando un cambio de la soberanía: Del Estado al individuo, del Estado a la corporación, de la ley al protocolo, de la institución a la red, (en definitiva) de la tierra a la nube. Este nuevo estrato de soberanía es «*The Stack*», la *Pila*, una mega estructura que funciona tanto como un sistema computacional como una arquitectura de gobierno. Para Bratton (2016: 19, 351), ésta urge no solo como un sistema técnico global, sino también como una geografía geopolítica; puede hacerlo porque también emerge del espacio político moderno y de su capacidad para ubicar, subdividir y ocupar nuevos mundos.

Esto confirma una vez más que la inteligencia artificial, sus recursos informativos y el mundo en línea constituyen una realidad total, que afecta completamente

la experiencia humana, tanto en sus proyecciones individuales y colectivas como en las económicas, políticas, privadas e institucionales.

III. INTELIGENCIA ARTIFICIAL, ALGORITMOS Y DECLINACIONES DE LA IGUALDAD

El tema de la igualdad está implicado en muy diferentes niveles por los desafíos legales que entraña la inteligencia artificial y entra en juego con toda su complejidad conceptual y la diversidad de sus proyecciones sustanciales. Por otra parte, hay que tener presente que el proyecto de la igualdad es un esquema abierto, en continua redefinición. El cambio de los contextos sociales, las posibilidades tecnológicas, las condiciones personales de vulnerabilidad, los riesgos discriminatorios, requieren nuevas herramientas de análisis y nuevas y más actualizadas configuraciones, siempre vinculadas a las cláusulas indeterminadas a través de las cuales se manifiesta el principio de igualdad (igual dignidad social, razonabilidad, igualdad sustancial, pleno desarrollo de la persona).

El primer segmento de intersección entre igualdad y desarrollo de los sistemas de inteligencia artificial se refiere a la posibilidad, no poco frecuente, de que estos sistemas y las evaluaciones algorítmicas puedan perpetuar (incluso indirecta o involuntariamente) modelos y esquemas discriminatorios, afectando de manera más severa y penalizante, a grupos que ya son minoritarios y marginalizados. Esto puede ser especialmente preocupante en ámbitos sensibles y delicados, en los que están en juego derechos fundamentales, como la justicia, la sanidad o el trabajo.

Muchos estudios han evidenciado que los *softwares* de previsión de delitos o de cálculo de la reincidencia pueden basarse en informaciones que acaban alimentando una especie de círculo vicioso, donde al final, lo que se criminaliza es la pobreza, la falta de educación, la criticidad de la situación familiar o incluso simplemente la ubicación de una persona en un determinado barrio (Umoja Noble: 2018). Y todo esto en un contexto de opacidad y escasa transparencia de las lógicas de funcionamiento de estas herramientas (Amato, 2020: 52).

En otros campos, además, también se han puesto de manifiesto resultados discriminatorios (objetiva e involuntariamente) o sesgados de otro modo. En el ámbito de la salud, por ejemplo, la construcción de conjuntos de datos con los que se entrenan los algoritmos utilizando esencialmente datos e información sanitaria relativos a personas blancas y de sexo masculino, crea no pocos problemas de coherencia y fiabilidad (así como de precisión) de las predicciones diagnósticas con referencia al estado clínico de las mujeres y/o personas de color.

Simoncini (2020: 167 ss.) habla abiertamente de un algoritmo estructuralmente inconstitucional, en cuanto está construido sobre un conjunto de datos que reflejan desde el principio elementos discriminatorios (racismo, sexismo) presentes en la representación informativa del mundo. El problema real, como

explica O'Neil (2017: 135 y 143), es entender «si hemos eliminado el prejuicio humano o simplemente lo hemos camuflado con la tecnología».

Evidentemente, se plantea un tema muy importante de formación ética (y de promoción de una cultura antidiscriminatoria, orientada a la transparencia y al pluralismo en la representación social) de los programadores y desarrolladores de sistemas o agentes de inteligencia artificial, para anticipar y reducir conflictos y/o prejuicios; y, al mismo tiempo, de formación tecnológica en las facultades jurídicas y en las oficinas de apoyo de los órganos institucionales y administrativos.

Hay una segunda clave de lectura de la relación entre la inteligencia artificial y el principio de igualdad. Los sistemas de inteligencia artificial serán uno de los grandes bloques del desarrollo económico mundial en este siglo, con repercusiones muy significativas en el mercado laboral, en la misma seguridad (y en muchos derechos y libertades) en los lugares de trabajo considerando las más frecuentes ocasiones de coexistencia e interacción entre humanos y agentes artificiales.

Categorías como la utilidad social u otras contenidas en los apartados 2 y 3 del art. 41 Const. italiana, y más en general todo el diseño constitucional de la sociedad, fundado en la centralidad del trabajo y en la igual dignidad social, deben adaptarse a los nuevos desafíos que nos esperan.

La revolución de la inteligencia artificial necesita ser acompañada y corregida por un pensamiento constitucional, debe producir una respuesta en términos de conceptualización de derechos y principios, del mismo modo que la revolución industrial produjo la evolución *welfarista* de los Estados liberales en el siglo XIX y el constitucionalismo social del siglo XX.

No es solo cuestión de puestos de trabajo que puedan perderse, y que en parte podrán ser compensados por nuevas necesidades y oportunidades laborales inducidas precisamente por la evolución tecnológica hacia la automatización.

Mantenerse al día con las transformaciones digitales requiere una actualización completa de los modelos de educación y formación profesional (Cingolani, 2019: 17 y 139), y un replanteamiento de algunas estructuras fundamentales del bienestar (buen vivir o *welfare*): vías de reinserción laboral, ingreso de ciudadanía o cualquier otro instrumento compensatorio para las situaciones de desempleo de larga duración. De lo contrario, la brecha social y económica vinculada precisamente al acceso y la familiaridad con las nuevas tecnologías corre el riesgo de amplificarse. Como señala Baricco (2018:320), «... el Juego es difícil. Tal vez divertido, pero demasiado difícil. (...) No dejaremos a nadie atrás no es una frase del Juego», y de transformarse en un factor de debilitamiento de la cohesión social y, en definitiva, de la democracia.

IV. INTELIGENCIA ARTIFICIAL Y JUSTICIA

En la versión actual del campo definido como *AI Law*, es ya un hecho que la decisión humana (en muchísimos sectores, como administración, trabajo,

hacienda, sanidad, justicia) pueda ser sustituida y/o acompañada/preparada por la decisión basada en procedimientos de tipo computacional. El tema es entender y establecer *qué y cuánto*, en cambio, consideramos que debe permanecer reservado a la esfera de la decisión humana. Uno de los sectores más estudiados es precisamente el relacionado con la aplicación de los sistemas de inteligencia artificial a la decisión jurisdiccional. La 'máquina' que juzga con rapidez y objetividad, que elimina de raíz retrasos en la administración de la justicia; pero que sobre todo no está condicionada por todos esos factores (subjetivos, emocionales, de adecuación profesional de las partes y del juez, incluso de prejuicios relacionados con el sexo o la orientación sexual, la raza, la religión, la nacionalidad, etc.) que hacen imperfecta la decisión humana. En *Criminal Sentences: Law Without Order*, un icónico ensayo de 1973, Frankel inició una serie de reflexiones sobre el *ruido* en la justicia penal que luego se desarrollaron en muchos estudios realizados sobre muestras estadísticas relevantes de sentencias penales. El resultado fue sorprendente, pero muy claro al evidenciar la dependencia del juicio final de factores (emocionales, raciales, relacionados incluso con el clima de ese día o con algunas pasiones deportivas de los Jueces) que tenían poco o nada que ver con la conducta, la actitud psicológica del reo, los otros elementos objetivos de la situación llevada a juicio. A la inversa, el uso de ordenadoras en el proceso podía representar una pieza importante hacia la afirmación del principio de legalidad penal como «cuerpo de reglas impersonales, aplicables de manera generalizada, vinculante para los jueces como para cualquier otra persona». En otras palabras, la decisión computacional sería objetiva, intrínsecamente imparcial, completamente desvinculada de los prejuicios que pertenecen a la experiencia humana. Con esta premisa teórica, la transición hacia una justicia digital parece casi irresistible. Ninguna de las objeciones que se pueden plantear parece definitiva, al menos no en todos los sectores o momentos procedimentales de la justicia. La justicia también está hecha de procedimientos repetitivos y estandarizados, y de decisiones meramente reiterativas de precedentes, donde la densidad interpretativa y el esfuerzo motivacional parecen muy bajos, casi inexistentes. En este sentido, creo que es un error abordar el problema en una óptica radical, de 'todo o nada'. Los sistemas de inteligencia artificial pueden aportar una contribución útil en muchas situaciones, seguramente en los procedimientos seriales y de bajo (o inexistente) contenido motivacional. Incluso en los casos más complejos, las indicaciones estadístico-predictivas del algoritmo pueden servir como base de conocimiento que puede ejercer una presión deflacionaria sobre el litigio (cuando, por ejemplo, pone en evidencia una clara y repetible orientación) o representar el punto de partida de una confrontación que puede desarrollarse en el mismo grado de juicio, o a través del uso de mecanismos de impugnación; siempre y cuando se respete el principio de no exclusividad (Zaccaria, 2023: 62, 65, 67). Sin embargo, no podemos ocultar los problemas. Desde mi punto de vista, motivación, impugnación (y obviamente la capacidad técnico-profesional del juez) siguen siendo las garantías constitucionales fundamentales del proceso y de la misma imparcialidad del Juez. La máquina

parece todavía demasiado opaca e incomprensible sobre por qué llega a una determinada síntesis, qué peso atribuye a los diferentes elementos y a la combinación entre ellos. En el *jus dicere* no solo es importante el resultado, o la rapidez (que también es una parte del 'justo' proceso); sino cómo se llega, a través de un enfrentamiento pleno y efectivo de las defensas (entre sí y con el Juez), donde la calidad de la decisión jurisdiccional sea también el reflejo de la capacidad del juez para hacer emerger la 'singularidad' de los hechos y los contextos humanos, para calibrar la decisión sobre ellos, en particular (o al menos) en cuestiones éticamente sensibles, o reivindicaciones inéditas, difíciles de colocar en una dimensión estandarizada. Precisamente, la motivación (que comprende hecho y derecho, y los conecta), es emblema de lo que sigue siendo «una empresa profundamente humana», ... (una) apuesta por la humanidad que implica una falibilidad» (Garapon, Lassègue, 2021: 252-253 y 278). No todo puede ser reducido a los carriles esquemáticos de la computación algorítmica. Los hechos que el Derecho está llamado a considerar presentan a veces una irreductible variedad, requieren razonabilidad, proporcionalidad; deben ser confrontados con principios abiertos y polisémicos; contienen y vierten sobre el Juez emociones, esperanzas, toda la complejidad de los problemas humanos. En 1881, en *The Common Law*, Oliver Wendell Holmes escribía que «*The life of the Law has not been logic: it has been experience*». Es el mismo riesgo que hoy corre la justicia si se reemplaza el juicio humano con el juicio tecnológico: perder el contacto con la vida real y con las personas. Esto se aplica especialmente a las decisiones que afectan directamente a las personas y su libertad. Me refiero al proceso penal, donde entran en juego, como tal vez en ningún otro subsistema jurisdiccional, derechos humanos fundamentales como la libertad personal, el derecho a la seguridad, la dignidad humana, el derecho a un juicio justo y a no ser considerados culpables hasta que se pruebe su responsabilidad, y el derecho del condenado a esperar la reinserción social. La justicia penal necesita de la 'mirada', del *intercambio* de empatía humana (con todas sus variables ruidosas) entre el juez y el acusado (o la víctima). En este campo, los automatismos pueden llevar a resultados irrazonables y desproporcionados. La Corte Suprema de Estados Unidos, en el caso *Woodson v. North Carolina* 428 U.S. 280 (1976), declaró inconstitucional la obligatoriedad de la pena de muerte para ciertos delitos, debido a que «trata a todas las personas condenadas por un delito designado no como seres humanos únicos e individuales, sino como miembros de una masa anónima e indiferenciada sometida al castigo ciego de la pena de muerte». Aparte de esto, los primeros experimentos de justicia 'predictiva' han demostrado, y quizás no debería sorprendernos dado que el algoritmo trabaja con 'nuestros' datos y entre 'nuestros' prejuicios, que los conjuntos de datos con los que estos sistemas están entrenados y de los que derivan predictivamente sus decisiones pueden ser incompletos, parciales, estar viciados por diversos *bias* o estar agregados de manera que produzcan resultados injustos, discriminatorios, inciertos o manipulados. En relación al caso Loomis (Nieva Fenoll, 2019: 59), ha destacado cómo algunos de los factores considerados en el

conjunto de datos de Compas (como: i) los antecedentes criminales en su familia y entre sus amigos; ii) la criminalidad en la zona de residencia; iii) el nivel educativo; iv) la situación financiera y laboral) pueden tener un impacto discriminatorio en términos raciales o socioeconómicos del reo. Y a menudo estas dos cosas se superponen. Hay un último punto crítico en la justicia algorítmica. Por definición, está construida sobre precedentes, en cierto sentido está *atrapada* en el pasado. Si bien es cierto que los precedentes son un elemento importante, a menudo decisivo, incluso en la forma humana de la jurisdicción, para la decisión algorítmica esta parece ser una condición genéticamente ligada a los modos de su formación. En cambio, el derecho *humano* está abierto al dinamismo, al descubrimiento de nuevos significados, a la confianza en que una opinión aislada puede llegar a ser compartida después de algunos años, frente a contextos culturales y sociales modificados. En esto radica la distancia fundamental entre la predictibilidad y la interpretación: lo primero tiene algo de estático, uniformador, «engrandece el presente» en una especie de «petrificación del tiempo» (Garapon, Lassègue, 2021: 184-185 y 92-93).

V. CONCLUSIONES

La inteligencia artificial es una tecnología transformadora. Estamos ante una nueva realidad. Se trata de una tecnología *generativa* que evoluciona también a la luz de cómo las personas trabajan con ella y a través de ella (Balkin, 2015: 47). No es casualidad que muchos académicos hablen de generaciones de inteligencia artificial, exactamente como nosotros, los juristas, hablamos de generaciones de derechos fundamentales. La adaptación sigue dinámicas fragmentadas. En algunos casos, podemos seguir trabajando con nuestros recursos conceptuales, las categorías y los principios tradicionales, simplemente redefiniendo sus contenidos e implicaciones en relación con un escenario claramente cambiado. En otros aspectos, las preguntas comienzan a ser más profundas, y se siente la sensación de un salto cualitativo y escalar en los problemas y reflexiones a los que estamos llamados a enfrentarnos. Estamos recorriendo caminos verdaderamente desconocidos, y no todo puede ser previsto y definido hoy. Sería un pecado de presunción, una simplificación innecesaria. Este *nuevo mundo* también se refleja en la formación del jurista, y más generalmente en la formación adecuada para las profesiones que deberán moverse alrededor y dentro de la evolución de la inteligencia artificial. La ética de la sociedad tecnológica necesita interdisciplinariedad (Coeckelbergh, 2020: 178,179), la capacidad de mantener una ventana abierta al mundo de los hechos científicos, que se han convertido en los factores que más inciden y determinan el desarrollo del estado-comunidad, de la conciencia social, de las dinámicas culturales; y creo que esto también debe ser válido en sentido contrario. La revolución digital será uno de los mayores acontecimientos de la historia humana. Nos desafiará en las grandes categorías de nuestro actuar, como

la inteligencia, la autonomía, la responsabilidad, la forma de socializar, pero en este desafío probablemente «nos invitará a reflexionar más seriamente y con menos complacencia sobre quiénes somos, podríamos ser o querríamos llegar a ser, y por lo tanto sobre nuestras responsabilidades y la comprensión que tenemos de nosotros mismos» (Floridi, 2022: 339). El Derecho, especialmente con sus principios fundamentales, tendrá un papel importante en este contexto en movimiento. Hablar de inteligencia artificial y Derecho ya no es una hipótesis de trabajo excéntrica, sino una empresa cultural, científica y práctica, ahora arraigada y desarrollándose de manera exponencial, atravesando sectores disciplinarios y obligándolos a enfrentarse con una dimensión material holística, total y al mismo tiempo fragmentada en una pluralidad de segmentos temáticos. Mucho dependerá de este trabajo coral para mantener y orientar el desarrollo de esta tecnología hacia la centralidad del ser humano, su experiencia y sus derechos.

BIBLIOGRAFÍA CITADA

- Amato S. (2020). *Biodiritto 4.0*, Giappichelli
- Balkin J. (2015). The Path of Robotics Law. *Californian Law Review*
- Baricco A. (2018). *The Game*, Einaudi
- Bratton, B.H. (2016). *The Stack: On Software and Sovereignty*, MIT Press
- Calasso, R. (2017). *L'innominabile attuale*, Adelphi
- Cardon, D. (2016), *Che cosa sognano gli algoritmi*, Mondadori
- Cingolani R. (2019). *L'altra specie. Otto domande su noi e loro*, IlMulino
- Crawford, K. (2021). *Né intelligente né artificiale. Il lato oscuro dell'IA*, IlMulino
- Floridi, L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore
- Coeckelbergh, M. (2020). *AI Ethics*, MIT Press
- D'Aloia, A. -Ed.- (2020). *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Franco Angeli.
- Foer, F. (2017). *I nuovi poteri forti*, Longanesi
- Frosini T.E. (2020). Il costituzionalismo della società tecnologica, *Diritto dell'informazione e dell'informatica*, 4
- Garapon, A., Lassègue, J. (2021). *La Giustizia digitale*, Il Mulino
- Hildebrandt M. (2015). *Smart technologies and the End(s) of Law. Novel Entanglements of Law and Technology*, Cheltenham
- Maestri E. (2017). Lex informatica e diritto. Pratiche sociali, sovranità e fonti nel cyberspazio, *Ars interpretandi*, 1
- O'Neil, C. (2017). *Armi di distruzione matematica*, Bompiani
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. The Belknap Press of Harvard University Press
- Resta G. (2019). Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza. *Politica del diritto*, 2

- Reichberg, G.M. e Syse, H. (2021). Applying AI on the Battlefield: The Ethical Debates, von Braun, J., Archer, M.S., Reichberg, G.M., Sánchez Sorondo, M. (dirs.), *Robotics, AI, and Humanity. Science, Ethics, and Policy*, Cham, Springer
- Sassi, S. (2021). *Disinformazione contro Costituzionalismo*, Editoriale Scientifica
- Simoncini, A. (2020). L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà. D'Aloia, A. (dir.), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, FrancoAngeli
- Umoja Noble S. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism*, NYU Press
- Zaccaria, G. (2023). *La responsabilità del giudice e l'algoritmo*, Mucchi Editore
- Zuboff, S. (2019). *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press

TITLE. *Artificial intelligence, fundamental rights, constitutional democracy*

ABSTRACT. AI (and in general the digital world) is already, and will become more and more, an essential element in people's lives and in the system of social relations; and one of the greatest events in human history. It will challenge us in the major categories of our actions, such as intelligence, autonomy, responsibility, the way we socialize, but in this challenge it will probably «invite us to reflect more seriously and with less complacency on who we are, could be or would like to become, and therefore on our responsibilities and the understanding we have of ourselves». Law, especially with its fundamental principles, will have an important role to play in this moving context, to maintain and guide the development of this technology towards the centrality of the human being, his experience and his rights.

RESUMEN La IA (y en general el mundo digital) ya es, y lo será cada vez más, un elemento esencial en la vida de las personas y en el sistema de relaciones sociales; y uno de los mayores eventos de la historia humana. Nos desafiará en las grandes categorías de nuestro actuar, como la inteligencia, la autonomía, la responsabilidad, la forma de socializar, pero en este desafío probablemente «nos invitará a reflexionar más seriamente y con menos complacencia sobre quiénes somos, podríamos ser o querríamos llegar a ser, y por lo tanto sobre nuestras responsabilidades y la comprensión que tenemos de nosotros mismos». El derecho, especialmente con sus principios fundamentales, tendrá un papel importante en este contexto en movimiento, para mantener y orientar el desarrollo de esta tecnología hacia la centralidad del ser humano, su experiencia y sus derechos.

KEY WORDS. *Artificial intelligence, Fundamental Rights, Constitutional democracy, bias, equal protection*

PALABRAS CLAVE. *Inteligencia artificial, Derechos fundamentales, Democracia constitucional, prejuicios y discriminación, igualdad de protección*

FECHA DE RECEPCIÓN: 10.01.2025

FECHA DE ACEPTACIÓN: 20.02.2025

CÓMO CITAR: D'Aloia, A. (2025). Inteligencia artificial, derechos fundamentales, democracia constitucional. *Teoría y Realidad Constitucional* 55, 101-114.