

INTELIGENCIA ARTIFICIAL Y DERECHOS HUMANOS

Rafael de Asís

MATERIALES DE FILOSOFÍA DEL DERECHO

Nº 2020 / 04

ISSN: 2531-0240

SEMINARIO PERMANENTE GREGORIO PECES-BARBA

GRUPO DE INVESTIGACIÓN

“Derechos humanos, Estado de Derecho y Democracia”

Serie: Materiales de Filosofía del Derecho

Número: 2020/04

ISSN: 2531-0240

Dirección de la serie: Rafael de Asís
Francisco Javier Ansuátegui

Editor: Seminario Gregorio Peces-Barba
Grupo de investigación “Derechos humanos, Estado de Derecho y Democracia”

Serie disponible en <http://hdl.handle.net/10016/24630>

Dirección: Seminario Gregorio Peces-Barba
Avd. de Gregorio Peces-Barba Martínez, 22
28270 Colmenarejo (Madrid)

Web: <http://www.seminario-gregorio-peces-barba.es>

Correo electrónico: info@seminario-gregorio-peces-barba.es



Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 España ([CC BY-NC-ND 3.0 ES](https://creativecommons.org/licenses/by-nc-nd/3.0/es/))

INTELIGENCIA ARTIFICIAL Y DERECHOS HUMANOS¹

Rafael de Asís²

rafael.asis@uc3m.es

Abstract: This paper describes the presence of human rights in the reflection on the use and scope of artificial intelligence. Artificial intelligence is presented as a tool that poses positive and negative challenges to the human rights discourse.

Keywords: Human Rights, Artificial Intelligence, Due Process, Reasoning.

Una cuestión actual

A finales de 2018, la Unión de Libertades Civiles de Nueva York, reveló que en 2017, el Software de Evaluación de Clasificación de Riesgos (RCA) que utilizaba desde 2013 el Servicio de Inmigración y Control de Aduanas de los Estados Unidos para ayudar a decidir, en procesos de deportación, si un inmigrante debía ser detenido o si podía ser puesto en libertad bajo fianza hasta el momento de la decisión definitiva, había sido manipulado para favorecer las detenciones.

El 24 de octubre de 2019, el periódico *El País*, publicaba que un equipo de investigadores había demostrado que un algoritmo usado para analizar los riesgos para la salud de millones de pacientes en EE.UU., discriminaba sistemáticamente a la población negra.

¹ Este texto tiene su origen en la conferencia que pronuncié en el Congreso sobre “Inteligencia Artificial y Derechos Humanos”, UNED, Madrid, 11 de noviembre de 2019.

² Catedrático de Filosofía del Derecho. Instituto de Derechos Humanos Bartolomé de las Casas. Universidad Carlos III de Madrid.

El 7 de noviembre de 2019, se celebró en Madrid, el *Foro Sanitas* sobre *El reto de la disrupción en medicina*, una de cuyas conclusiones fue que la inteligencia artificial nos permitirá hacer una medicina más humana.

Numerosos estudios han demostrado que existen empresas que discriminan en materia de precios a través de la Inteligencia artificial y cómo las normas antidiscriminación no son eficaces contra esta práctica³.

Un estudio publicado por el Berkman Klein Center de la Universidad de Harvard en septiembre de 2018, sobre inteligencia artificial y derechos humanos ha destacado la proyección de la inteligencia artificial en aspectos muy diferentes como pueden ser los procesos de préstamo, la salud, la contratación, la información...⁴.

Ante esta realidad no cabe considerar que el uso de la inteligencia artificial no deba regularse o que esta regulación deba quedar en manos de las compañías tecnológicas.

¿Cómo regular?

Una sola aplicación de inteligencia artificial puede impactar en una gran cantidad de derechos. En este sentido, hace cuatro años publicaba *Una mirada a la robótica desde los derechos humanos*⁵, destacando la necesidad de alcanzar un consenso en relación con los referentes éticos desde los que afrontar la regulación de la inteligencia artificial y la robótica.

Pues bien, recientemente, un estudio publicado por la Escuela Politécnica Federal de Zurich⁶, llega a la conclusión de que existe un acuerdo global en torno a cinco

³ Vid. Zuiderveen Borgesius, F., “Algorithmic decision-making, price discrimination, and European non-discrimination law”, disponible en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413556 Consultado el 20 de septiembre de 2019.

⁴ *Artificial Intelligence & Human Rights: Opportunities & Risks*, septiembre de 2018, Filippo A. Raso, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, Levin Kim, The Berkman Klein Center for Internet & Society. Universidad de Harvard.

⁵ Dykinson, Madrid 2015.

⁶ *Artificial Intelligence: the global landscape of ethics guidelines*, Health Ethics & Policy Lab, ETH Zurich, 2019.

principios éticos: transparencia, justicia y equidad, no maleficencia, responsabilidad y privacidad. Ahora bien, en el estudio, se subraya que existen desacuerdos sustanciales en relación con la forma en que se interpretan estos principios, por qué se consideran importantes y cómo deben implementarse.

A la vista de ello cabe preguntarse si existe realmente dicho consenso. Y es que, como es sabido, en el campo filosófico existen dos grandes posturas (con muchas variantes en su interior) sobre como enfrentarse a las nuevas tecnologías. Por un lado están los bioconservadores, que subrayan las consecuencias desastrosas que los avances tecnológicos pueden producir a la humanidad y al medio ambiente, y por otro los transhumanistas, que pretende emplear la tecnología para mejorar la vida de las personas y para resolver los problemas sociales contemporáneos. Las diferencias entre unos y otros son sensibles y se proyectan tanto en el ámbitos epistemológico, como en el ontológico y en el deontológico⁷.

Se habla de ausencia de referentes éticos para hacer frente a los problemas sociales que son consecuencia de la evolución de la ciencia y la tecnología. Ahora bien, a diferencia de lo que ha ocurrido en otros momentos de la historia de la humanidad en donde se han producido grandes cambios tecnológicos, en la actualidad estamos mejor situados, ya que contamos con unos instrumentos a los que hemos otorgado un cierto valor universal: los derechos humanos. Así, a pesar de que existan diferentes maneras de entender los derechos humanos que se manifiestan en distintas teorías, y a pesar de que no podamos pensar que exista un acuerdo final sobre su significado⁸, contar con los derechos implica establecer unos puntos de partida muy importantes en la reflexión.

Obviamente, tomar como referencia los derechos no va a servir para solucionar todos los problemas. Pero si que puede servirnos para reorientar algunas de las discusiones.

⁷ Al respecto vid. Llano Alonso, F., *Homo Excelsior. Los Límites Ético Jurídicos del Transhumanismo*, Tirant Lo Blanch, Valencia 2018. Vid. recientemente del mismo autor: “Transhumanismo, vulnerabilidad y dignidad humana”, en *Deusto Journal of Human Rights*, 4, 2019, pp. 39 y ss.

⁸ Al respecto resulta significativa la anécdota que cuenta J. Maritain en su obra *El hombre y el Estado*, (trad. de J.M. Palacios, Encuentro, Madrid 2002, p.84): “Durante una de las reuniones de la Comisión nacional francesa de la UNESCO en que se discutía sobre los derechos del hombre, alguien manifestó su extrañeza al ver que ciertos defensores de ideologías violentamente opuestas se habían puesto de acuerdo para redactar una lista de derechos. Claro - replicaron ellos- estamos de acuerdo en esos derechos a condición de que no se nos pregunte por qué”.

Por ejemplo, en el campo de la discapacidad y las nuevas tecnologías, la aproximación basada en derechos llamaría la atención sobre la cuestión de las barreras sociales y la discriminación.

En efecto, en este campo, cuando se hace alusión a la incidencia de las nuevas tecnologías suelen primar dos tipos de reflexiones. Por un lado, la de aquellos que presentan los avances tecnológicos como remedios que consiguen reducir los efectos de las deficiencias o incluso, terminar con ellos. Por otro lado, la de quienes enfatizan como las nuevas tecnologías pueden ser un instrumento más de discriminación hacia estas personas en términos de accesibilidad. Pues bien, en la mayoría de los casos, estas dos reflexiones se hacen desde un enfoque de la discapacidad, centrado en la condición de la persona, esto es en la “deficiencia”, que no es el de los derechos. El enfoque de derechos es el que considera a la discapacidad como una situación fruto de barreras que son, en la mayoría de los casos, de índole social⁹. Por eso, desde este enfoque debería ser prioritario la lucha contra las barreras y no la lucha contra las llamadas deficiencias, al igual que debería priorizar esta lucha frente a los programas de mejora.

Los derechos humanos ya están aquí

La referencia de los derechos en la discusión sobre cómo regular la inteligencia artificial es un hecho.

En mayo de 2018, después de discusiones y reuniones entre expertos de diferentes campos del conocimiento, se aprobó la *Declaración de Toronto* (firmada por Amnistía Internacional, Access Now, el Observatorio de Derechos Humanos y la Fundación Wikipedia, entre otros). Se trata de un documento con tres secciones principales. En la primera, se establece el deber de los Estados de prevenir la discriminación en el diseño o la implementación de los sistemas de aprendizaje automático en contextos públicos o mediante asociaciones público-privadas. En la segunda sección se describe las responsabilidades de los actores privados en el contexto del desarrollo y la implementación de sistemas de inteligencia artificial. Y en la tercera sección se afirma el derecho a un recurso efectivo y a llamar a cuentas a los responsables de las violaciones.

⁹ Vid. mi libro *Sobre discapacidad y derechos*, Dykinson, Madrid 2013.

La Declaración de Toronto busca proteger los derechos de todos los individuos y grupos y promover la diversidad y la inclusión. Subraya que la inclusión, la diversidad y la equidad, son componentes clave para garantizar que los sistemas de aprendizaje automático no creen ni perpetúen la discriminación, en particular en relación con grupos marginados¹⁰.

En octubre de 2018, la organización *The Public Voice*, aprobó las *Directrices Universales para la Inteligencia Artificial*, documento respaldado por 50 organizaciones científicas y más de 200 expertos de todo el mundo, que tiene como objetivo “maximizar los beneficios de la IA, minimizar el riesgo y garantizar la protección de los derechos humanos”¹¹. En este mismo mes, la Conferencia Internacional de instituciones dedicadas a la protección de datos y la privacidad (ICDPPC), adoptó la *Declaración sobre ética y protección de datos en la Inteligencia Artificial*, en la que se afirma que “los derechos humanos deben ser el marco desde el que afrontar los retos éticos de las nuevas tecnologías. Y los responsables de que esto sea así son, en primer lugar, los Estados y las organizaciones internacionales como Naciones Unidas. Pero esa responsabilidad nos compete a todos y, como no, a las empresas e industrias”¹².

En el ámbito de la Unión Europea se han elaborado numerosos documentos en esta línea. Entre todos destaca el Comunicado de 8 de abril de 2019 de la Comisión Europea, *Siete requisitos esenciales para lograr una inteligencia artificial fiable*. Estos siete requisitos son: (i) la intervención y supervisión humanas (los sistemas de inteligencia artificial deben facilitar sociedades equitativas, apoyando la intervención humana y los derechos fundamentales, y no disminuir, limitar o desorientar la autonomía humana); (ii) la robustez y seguridad (los algoritmos deben ser suficientemente seguros, fiables y sólidos para resolver errores o incoherencias durante todas las fases del ciclo de vida útil de los sistemas de inteligencia artificial); (iii) privacidad y gestión de datos (los ciudadanos deben tener pleno control sobre sus propios datos, al tiempo que los datos que les conciernen no deben utilizarse para perjudicarles o discriminarles); (iv) transparencia

¹⁰ Disponible en: https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf (última consulta el 12 de noviembre de 2019).

¹¹ Disponible en <https://thepublicvoice.org/ai-universal-guidelines/> (última consulta el 12 de noviembre de 2019).

¹² Disponible en https://apdc.gencat.cat/web/.content/04-actualitat/noticies/documents/ICDPPC-40th_AI-Declaration_ADOPTED.pdf (última consulta el 12 de noviembre de 2019).

(debe garantizarse la trazabilidad de los sistemas de inteligencia artificial); (v) diversidad, no discriminación y equidad (los sistemas de inteligencia artificial deben tener en cuenta el conjunto de capacidades, competencias y necesidades humanas, y garantizar la accesibilidad); (vi) bienestar social y medioambiental (los sistemas de inteligencia artificial deben utilizarse para mejorar el cambio social positivo y aumentar la sostenibilidad y la responsabilidad ecológicas); (vii) rendición de cuentas (deben implantarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de inteligencia artificial y de sus resultados)¹³.

Por su parte, el Consejo de Europa también ha aprobado declaraciones en esta línea. Destacaré tres: las *Directrices sobre Inteligencia Artificial y Protección de Datos* de enero de 2019, la *Declaración del Comité de Ministros sobre las capacidades manipuladoras de los procesos algorítmicos* de febrero de 2019 y *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* de mayo de 2019.

Las *Directrices sobre Inteligencia Artificial y Protección de Datos* comienzan señalando: “La protección de la dignidad humana y la salvaguarda de los derechos humanos y las libertades fundamentales, en particular el derecho a la protección de los datos personales, son esenciales al desarrollar y adoptar aplicaciones de IA que puedan tener consecuencias para las personas y la sociedad. Esto es especialmente importante cuando las aplicaciones de IA se utilizan en los procesos de toma de decisiones”¹⁴.

Por su parte, en la *Declaración del Comité de Ministros sobre las capacidades manipuladoras de los procesos algorítmicos* de febrero de 2019, puede leerse: “Los niveles finos, subconscientes y personalizados de persuasión algorítmica pueden tener efectos significativos en la autonomía cognitiva de los individuos y su derecho a formar opiniones y tomar decisiones independientes. Estos efectos permanecen poco explorados pero no se pueden subestimar. No solo pueden debilitar el ejercicio y el disfrute de los derechos humanos individuales, sino que pueden conducir a la corrosión del fundamento mismo del Consejo de Europa. Los pilares centrales de los derechos humanos, la democracia y el estado de derecho se basan en la creencia fundamental en la igualdad y

¹³ Disponibles en <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (última consulta el 12 de noviembre de 2019).

¹⁴ Disponibles en <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8> (última consulta el 13 de noviembre de 2019).

la dignidad de todos los humanos como agentes morales independientes”. En esta Declaración se “destaca el papel social de la academia en la producción de investigaciones y consejos independientes, basados en evidencias e interdisciplinarios, dirigidos a los agentes que toman decisiones, sobre la capacidad de las herramientas algorítmicas para mejorar o interferir en la soberanía cognitiva de los individuos. Esta investigación debe tener en cuenta la diversidad existente en las sociedades, y debe incluir todos los antecedentes y edades de los usuarios, no solo con respecto a sus comportamientos como consumidores, sino teniendo también en cuenta impactos más amplios en su bienestar emocional y elecciones personales en contextos sociales, institucionales y políticos”¹⁵.

Por último, en *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* de mayo de 2019, se destaca la necesidad de realizar evaluaciones de impacto en derechos humanos en relación con los sistemas de inteligencia artificial¹⁶.

Esta presencia de los derechos humanos en los documentos que se pronuncian sobre la regulación de la inteligencia artificial está presente también fuera de Occidente. Son significativos al respecto, los denominados *Principios de Pekín*, aprobados en mayo de 2019, por la Academia de Inteligencia Artificial de Pekín (AIAB), la Universidad de Pekín, la Universidad de Tsinghua, el Instituto de Automatización, el Instituto de Tecnología de Computación de la Academia de Ciencias de China, y una liga industrial de inteligencia artificial formada por firmas como Baidu, Alibaba y Tencent. En estos principios se habla del desarrollo sostenible, de la privacidad, de la dignidad, de la libertad, de la autonomía y de los derechos, como referentes para la investigación y el desarrollo de la inteligencia artificial. Así, se afirma que “el desarrollo de la IA debe reflejar la diversidad y la inclusión, y debe diseñarse para beneficiar a la mayor cantidad de personas posible, especialmente a aquellos que, de lo contrario, serían fácilmente descuidados o insuficientemente representado”¹⁷.

¹⁵ Disponible en https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b (última consulta el 13 de noviembre de 2019).

¹⁶ Disponible en <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64> (última consulta el 13 de noviembre de 2019).

¹⁷ Disponibles en <https://www.baai.ac.cn/blog/beijing-ai-principles> (consultado por última vez el 11 de noviembre de 2019)

La utilización de los derechos humanos como referentes para la regulación de las nuevas tecnologías, que ya era una constante en el ámbito biomédico, es un hecho en el campo de la inteligencia artificial, y se extiende a otros campos como el de la neurociencia¹⁸.

En nuestro país, dentro de esta tendencia es significativo el reconocimiento de los llamados derechos digitales, que se realiza en los artículos 80 a 97 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Para un estudio de los derechos humanos resulta de especial interés el análisis de estos derechos y su caracterización general, con el objetivo de encontrar su lugar en la Teoría General de los Derechos. Se trata de una tarea sobre la que existen ya algunos estudios¹⁹ y que plantea retos interesantes para la Teoría del Derecho, el Derecho Constitucional y el Derecho Internacional, entre otras disciplinas.

Dos de estos retos, que tienen que ver con rasgos de los derechos, coinciden con dos de los déficits más importantes que tiene la satisfacción de los derechos humanos en el mundo contemporáneo: su incidencia en el ámbito de las relaciones privadas y su necesaria protección a nivel internacional.

Buena parte del campo de actuación de los derechos digitales se desenvuelve en las relaciones entre empresas y ciudadanos. La satisfacción de los derechos digitales requiere, en muchos casos, de la complicidad de grandes y pequeñas compañías privadas. Sin embargo, todavía nos cuesta aceptar que los derechos humanos, en general, poseen vigencia en ese plano. Es cierto que teóricamente se trata de un tema resuelto y que existen

¹⁸ En este sentido Rafael Yuste, del Centro de Neurotecnología de la Universidad de Columbia en Nueva York, impulsor del proyecto BRAIN, se refiere a 5 neuroderechos humanos (la privacidad mental, la identidad personal, el libre albedrío, el acceso equitativo, la no discriminación (Vid. la NeuroRights Initiative, <https://www.ntc.columbia.edu/neurorights-initiative/>). También Marcello Ienca y Roberto Andorno (“Towards new human rights in the age of neuroscience and neurotechnology”, en *Life Sciences, Society and Policy*, 2017, disponible en <https://lssjournal.biomedcentral.com/track/pdf/10.1186/s40504-017-0050-1>) se han referido al derecho a la libertad cognitiva (decidir si desea utilizar una neurotecnología determinada o rechazarla); la privacidad mental (aislar los datos de su cerebro o compartirlos públicamente); la integridad mental (no ser dañado física o psicológicamente por la neurotecnología; la continuidad psicológica (estar protegido de las alteraciones de su sentido de identidad que no haya autorizado).

¹⁹ Vid. por ejemplo Rebollo Delgado, L. y Zapatero Martín, P., *Derechos Digitales*, UNED-Dykinson, 2019.

grandes documentos internacionales al respecto (como los “Principios Rectores sobre las empresas y los derechos humanos” de Naciones Unidas). Pero otra cosa es la práctica y la garantía de esos principios.

Por otro lado, la globalidad de la sociedad digital hace que una eficaz protección de los derechos digitales, requiera de la utilización de sistemas de protección internacional. No basta con la protección nacional ni incluso, en nuestro caso, con la europea. Sin embargo, todavía nos cuesta aceptar el papel de los sistemas de protección internacional de los derechos. Dudamos tanto del acuerdo internacional, del que en muchas ocasiones se salen las grandes potencias, como de la legitimidad de las instituciones internacionales.

Y estos dos problemas teóricos con indudable repercusión práctica, nos alertan de las dificultades que nos vamos a encontrar para satisfacer estos derechos y nos confirman la necesidad de llevar a cabo cambios sociales que permitan considerar, por un lado, que los derechos son los referentes también de las relaciones privadas y económicas, y por otro, que los sistemas de protección internacional deben ser eficaces y tomados en serio.

Volviendo a los documentos y textos normativos sobre derechos e inteligencia artificial, conviene advertir que a pesar de esta proliferación, la realidad es que la inteligencia artificial sigue estando separada de los derechos. Los departamentos de las empresas que la utilizan no atienden a su impacto sobre los derechos y dejan todo ello a las indicaciones de una autoridad reguladora que, en muchas ocasiones, ni existe ni tiene competencias o medios para establecer un control.

Riesgo y oportunidad

Un rasgo común a todo ese movimiento de incorporación de los derechos al que me he referido, es la adopción de un punto de vista que se centra en los aspectos negativos de estas tecnologías en la garantía y satisfacción de los derechos²⁰. Normalmente este es el enfoque que prima cuando se toma como referencia los derechos. Y es cierto que el

²⁰ Yo también me he referido a estos problemas en *Una mirada a la robótica desde los derechos humanos*, cit., pp. 50 y ss.

uso de la inteligencia artificial presenta riesgos relevantes sobre todo para las personas en situación de vulnerabilidad.

Así, por ejemplo, volviendo a la discapacidad, su tratamiento común y mayoritario, centrado todavía en un modelo médico, presenta una serie de rasgos que, si no son advertidos en la aplicación de la inteligencia artificial, pueden agravar la situación de discriminación de las personas con discapacidad. No hay que pasar por alto que las aplicaciones de inteligencia artificial trabajan en función de los datos y la información que se les proporcionan y que si no se corrigen, reproducen lógicas discriminatorias.

La consideración de la discapacidad como una enfermedad, como algo maligno que debe ser corregido, o como algo especial, fuera de lo normal, son ejemplos de esos rasgos que pueden provocar sesgos en la aplicación de la inteligencia artificial. También es importante tener en cuenta como todavía, en muchos ámbitos sociales, la discapacidad sigue siendo invisible lo que sin duda puede incidir en el uso de esta tecnología.

Pero tal vez, el peligro más relevante venga de la mano de la problemática relación entre inteligencia artificial y diversidad.

Esta problemática relación ha sido estudiada en relación con el género y la raza, subrayando sesgos presentes en aplicaciones de inteligencia artificial que discriminan a mujeres, personas de color, minorías sexuales.... En este ámbito destaca el informe del Instituto *AI Now* de la Universidad de Nueva York, *Discriminating Systems. Gender, Race, and Power in AI* de abril de 2019²¹.

Ese mismo Instituto, convocó el 28 de marzo de 2019, junto al Centro de Estudios de Discapacidad de esa misma Universidad y Microsoft, a especialistas en discapacidad, programadores de Inteligencia Artificial e investigadores de informática para reflexionar sobre discapacidad, sesgo e Inteligencia Artificial. Como resultado de esa reunión *AI Now* publicó, en noviembre de 2019, otro informe, *Disability, Bias, and AI*²², en el que aparecen interesantes reflexiones y ejemplos sobre posibles sesgos del uso de la Inteligencia Artificial en relación con las personas con discapacidad.

²¹ Disponible en <https://ainowinstitute.org/discriminatingsystems.pdf>. (última consulta el 12 de noviembre de 2019).

²² Dponible en <https://ainowinstitute.org/disabilitybiasai-2019.pdf>. (última consulta el 12 de noviembre de 2019).

Analizar como es posible hacer que esos sesgos no se produzcan resulta verdaderamente importante en la lucha contra la discriminación de las personas con discapacidad. Y para ello, la formación de todos los agentes implicados en los derechos de las personas con discapacidad y, por tanto, en el tratamiento de la discapacidad desde un enfoque de derechos humanos, parece imprescindible. Como también lo parece el establecimiento de canales de diálogo permanentes entre la ética y la tecnología. Ahora bien, en todo ello, resulta esencial la participación de las propias personas con discapacidad: el cumplimiento del lema “Nada sobre nosotros sin nosotros” debe ser un requisito ineludible.

En todo caso, resulta necesario analizar también el lado positivo de la utilización de la inteligencia artificial; contemplar la inteligencia artificial como una oportunidad para los derechos.

Un ejemplo de esta posibilidad viene de la mano del derecho al debido proceso. Como es sabido, cuando hablamos de debido proceso estamos haciendo referencia a una serie de rasgos que deben ser respetados en todos y cada uno de los procesos y que en ocasiones se ha concretado en el derecho a la tutela judicial efectiva. Su primera manifestación se encuentra en la Carta Magna de 1215. En la actualidad, el debido proceso posee un contenido complejo que integra: el derecho de acción o de acceso a la jurisdicción, el derecho a un proceso sin dilaciones indebidas, el derecho a un proceso público, la presunción de inocencia, el derecho a ser informado de la acusación, el derecho a la intervención del intérprete, el sometimiento del juzgador al Derecho, la independencia del juzgador, la imparcialidad del juzgador, el derecho a la firmeza de las decisiones (cosa juzgada), el derecho al *non bis in idem*, el derecho a la ejecución de la sentencia, el derecho al juez natural, la exigencia de motivación suficiente, la prohibición de indefensión, el derecho a la prueba, el derecho a todas las garantías, el derecho a la asistencia letrada, el derecho a no declarar contra sí mismo, el derecho a no confesarse culpable, el derecho a los recursos y el derecho a una motivación completa²³.

Pues bien, en 2017, el Presidente del Tribunal Supremo de los EEUU, John Roberts, visitó el Instituto Politécnico de Rensselaer (Troy, Nueva York), que es uno de

²³ Vid. De Asís, R., “Notas sobre el debido proceso”, en *Papeles de Teoría y Filosofía del Derecho*, n. 6, 2010. Disponible en https://e-archivo.uc3m.es/bitstream/handle/10016/9224/notas_sobre_el_debido_proceso_rafael_de_asis_%202010.pdf?sequence=1&isAllowed=y.

los centros de docencia e investigación en ciencia e ingeniería más importantes de EE.UU, y se le hizo una pregunta: ¿Se puede prever un día en el que las máquinas inteligentes, impulsadas con inteligencias artificiales, ayuden en la determinación de hechos en los tribunales e incluso en las decisiones de calificación? Y su respuesta fue: "Es un día que está aquí...y está poniendo una presión significativa en cómo el poder judicial se encarga de hacer las cosas".

Ciertamente, en el contexto jurídico asistimos a un momento en el que la Inteligencia artificial está empezando a implementarse de manera progresiva, sobre todo en lo referente a sistemas de finalidad predictiva²⁴. En 2016, el University College de Londres, presentó un software de aprendizaje automático capaz de presentar los mismos veredictos que los jueces. Para ello realizó un estudio en el que se analizaron 584 decisiones del Tribunal Europeo de Derechos Humanos en asuntos relacionados con los artículos 3, 6 y 8 de la Convención, aplicando un algoritmo con el objetivo de encontrar patrones en el texto. La finalidad última consistía en ver si el *software* podía predecir el fallo. Y en un 79% de los casos lo consiguió²⁵.

La utilización de la inteligencia artificial en la ámbito de la decisión judicial es una cuestión polémica. Para algunas personas ese uso es, por la especificidad de la toma de decisiones en este campo, y debe ser, por los derechos en juego, muy limitado. No obstante, no hay que pasar por alto que buena parte de esta actuación consiste en gestionar datos e información²⁶.

Por otro lado, y aunque la mayoría de las aplicaciones actuales se basan en grandes datos y estadística²⁷, se han anunciado proyectos que intentan replicar la argumentación jurídica desde una inteligencia artificial que comprende el contexto²⁸, e incluso no hace

²⁴ Vid. al respecto, Solar Cayón, J.I., *La inteligencia artificial jurídica*, Aranzadi, 2019.

²⁵ Vid. Aletras, N., Tsarapatsanis, D., Preotjiuc-Pietro, D. y Lamos, V., "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective", en *PeerJ Comput. Sci.* 2:e93, 2016. Disponible en <https://peerj.com/articles/cs-93.pdf> (última consulta 13 de noviembre de 2019).

²⁶ Vid. Nieva Fenoll, J., *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid 2018, pp. 60 y ss.

²⁷ Vid. López Rincón, D., "Robots y abogacía", en Barrio, M. (dir.), *Derecho de los Robots*, Wolters Kluwer, Madrid 2019, pp. 229 y ss.

²⁸ Vid. <https://magiquo.com>

mucho podíamos leer en diferentes medios que el Ministerio de Justicia de Estonia estaba pensando en una inteligencia artificial que actúe a modo de juez, para encargarse de aquellos casos más sencillos. Si bien se aclaraba que la decisión final la tomaría siempre un juez humano.

A pesar de que en la argumentación jurídica, la relevancia de la interpretación, tanto de normas como de hechos, provoca que la discrecionalidad sea un factor de indudable peso, existen reglas que condicionan y validan el uso de argumentos y que pueden ser utilizadas en la aplicación de la inteligencia artificial a este campo²⁹.

El caso más conocido de aplicación de la inteligencia artificial al Derecho es el llamado caso Loomis. Se trata de un caso que tiene que ver, de nuevo, con la evaluación de riesgos.

Eric Loomis era un ciudadano que cuando conducía un vehículo, al parecer robado, se saltó un control de la policía siendo posteriormente detenido. Sus datos se introdujeron en una herramienta de evaluación de riesgos (COMPAS), que emitió un informe en el que se predecía un alto riesgo de reincidencia. Fue así condenado a una pena de prisión descartando el juez la libertad condicional apoyándose para ello en dicha herramienta. Loomis apeló ante la Corte Suprema de su Estado alegando la violación de su derecho al debido proceso por no poder cuestionar la fuerza probatoria de la herramienta al estar patentada. La Corte Suprema falló en contra de Loomis, determinando que no se violaba el debido proceso, con el argumento de que la evaluación de riesgo de COMPAS, no fue determinante para la decisión sobre él. La herramienta había servido como apoyo y para proporcionar al tribunal la mayor información posible. Loomis apeló ante la Corte Suprema de los EE. UU., que en junio de 2017 rechazó conocer su caso³⁰.

Casos como el de Loomis, han suscitado un importante debate en el campo jurídico y han hecho saltar las alarmas sobre la posible violación de derechos cuando se utiliza la inteligencia artificial en el campo de la justicia.

²⁹ Vid. al respecto De Asís, R., “Robótica, Inteligencia Artificial y Derecho”, en *Revista de Privacidad y Derecho Digital*, vol. 3, n. 10, 2018, pp. 27 y ss.

³⁰ Vid. *State v. Loomis*, en <https://harvardlawreview.org/2017/03/state-v-loomis/> (última consulta el 6 de diciembre de 2019).

Así, se ha hablado de la deshumanización de la jurisdicción, con una especial proyección en la atención a la víctimas (que ha sido una de las tendencias del Derecho penal y procesal contemporáneo), de los riesgos a una excesiva estandarización en las sentencias, del cuestionamiento de las decisiones de los Tribunales supremos que finalmente se producirá, de la falta de publicidad y garantías, de la afectación del derecho a un juicio justo, de la desigualdad de armas... En definitiva, parecería que el uso de la inteligencia artificial puede conllevar la violación del derecho al debido proceso.

Sin embargo, algunas de las críticas anteriores chocan con la manera con las que en muchas ocasiones describimos el Derecho, la actividad de los jueces o la argumentación jurídica en general.

Por otro lado, en el ámbito del proceso, la intervención de la tecnología a través de peritos, o el uso de la informática y la probabilidad es algo habitual. El uso de la tecnología del caso Loomis no es algo extraño (otra cosa puede ser la solución). Así, en el campo de la argumentación se analiza el uso de la estadística bayesiana y de la probabilidad en la resolución de determinados asuntos³¹. Es célebre en este sentido el conocido caso *People vs Collins* de la Corte Suprema de California³². En él, una anciana es asaltada por la calle por unos desconocidos, de los cuales sólo existen dos testimonios: el de la anciana que dice haber visto a una joven rubia huir, y el de un testigo que dice haber visto a una joven blanca con los cabellos rubios y coleta huir en un coche amarillo conducido por un joven negro con barba y gafas. Poco después la policía detuvo a una pareja que respondía a esa descripción. En el proceso, la acusación llamó a un experto en estadística para que, asumiendo que el hecho había sido cometido por una joven blanca con cabellos rubios y coleta y que había huido en un coche amarillo conducido por un joven negro con barba y gafas, examinara la probabilidad real de que los detenidos fueran los autores del hecho. Se individualizaron seis características relevantes (coche amarillo, hombre con gafas, mujer con coleta, mujer con cabellos rubios, hombre negro con barba, pareja de hombre negro y mujer blanca en un coche) y los valores de la probabilidad de cada característica se multiplicaron. Así se llegó al resultado de que una de cada doce millones de parejas tenían esas características, concluyendo en que existía una posibilidad sobre doce millones de que los detenidos, que tenían todas las características, fueran

³¹ Me he referido a ello en De Asís, R., *Sobre el razonamiento judicial*, McGraw-Hill, Madrid 1998.

³² Recogido en *California Reporter*, 1966, vol. 66, pp. 497 y ss. Vid. sobre este caso: Alonso García, E., *La interpretación de la Constitución*, Centro de Estudios Constitucionales, Madrid 1984, p. 543.

inocentes. La Corte de los Ángeles condenó basándose en ese estudio. Pero la Corte Suprema corrigió la decisión, basándose principalmente en los dos siguientes motivos. En primer lugar, afirmando que no había prueba que confirmara cada una de las probabilidades individuales ni que los seis factores fueran estadísticamente independientes³³. En segundo lugar, afirmando que ese tipo de análisis había apartado al tribunal de su verdadera función de valorar la prueba, existiendo la posibilidad de que hubiera alguna pareja con esas características en Los Ángeles o incluso que los testimonios fueran falsos.

El caso suele citarse como ejemplo para apartar el uso de la probabilidad del Derecho³⁴. Ahora bien, es importante darse cuenta de que ninguna de las razones que utiliza el Tribunal invalida el apoyo en este tipo de situaciones, no tanto como forma de decisión automática sino más bien como un instrumento más de la valoración.

Y en este sentido, la inteligencia artificial puede ser un instrumento más de apoyo en la argumentación y la motivación³⁵, que incluso sirva para su mejora sin menoscabo del debido proceso.

Ciertamente, el peligro mayor del uso de la inteligencia artificial en este ámbito puede venir de la mano de los posibles sesgos. Ahora bien, hay que ser conscientes de que los sesgos existen con y sin inteligencia artificial. Es más, gracias a ella puede ser posible identificar estos sesgos y eliminarlos³⁶. Eso sí, para ello debe exigirse el mayor grado de transparencia posible a la hora de determinar su uso y alcance.

³³ En efecto, la Corte de Los Ángeles, aplicó una regla del cálculo de probabilidades referida a sucesos independientes. En este ámbito se afirma que dos sucesos son independientes cuando la ocurrencia de uno de ellos no altera la probabilidad del otro. Así, la regla dice: la probabilidad de que ocurran simultáneamente dos sucesos independientes es igual al producto de la probabilidad de dichos sucesos. El problema, obviamente, radica en la determinación de cuándo los sucesos son independientes, ya que si no lo son, varía la regla de probabilidad a emplear.

³⁴ Normalmente en un tono crítico señalando insuficiencias que tal vez la inteligencia artificial pueda solucionar. Vid. sobre el tema Taruffo, M., “Judicial Decision and Artificial Intelligence”, en *Artificial Intelligence and Law*, 6, 1998. También Vera Sánchez, J.S., “La discrecionalidad judicial en dos enfoques reglados de la valoración de la prueba penal”, *Revista Jurídica de la Universidad Autónoma de Madrid*, 30, 2014-II, pp. 247 y ss.

³⁵ Vid. Nieva Fenoll, J., *Inteligencia artificial y proceso judicial*, cit., p. 71.

³⁶ Vid. Chen, D.L., “Judicial analytics and the great transformation of American Law”, en *Artificial Intelligence and Law*, 27, 2019, pp. 15 y ss.

En todo caso, al igual que ocurría en el plano general de la inteligencia artificial, se están produciendo ya declaraciones sobre su uso en relación con el proceso debido. Merece la pena citar en este sentido la *Carta Ética europea sobre el uso y desarrollo de la inteligencia artificial en el sistema judicial*, de la Comisión Europea para la Eficiencia de la Justicia, de diciembre de 2018, en la que se destacan cinco grandes principios: a) el respeto de los derechos fundamentales; b) la no discriminación; c) la calidad y seguridad; d) la transparencia, imparcialidad y equidad; y, e) el principio "bajo el control del usuario". En virtud de este último, "la autonomía del usuario debe aumentarse y no restringirse mediante el uso de herramientas y servicios de inteligencia artificial. Los profesionales del sistema de justicia deberían, en cualquier momento, poder revisar las decisiones judiciales y los datos utilizados para producir un resultado y continuar sin estar obligados a ello a la luz de las características específicas de ese caso en particular. El usuario debe ser informado en un lenguaje claro y comprensible sobre si las soluciones ofrecidas por las herramientas de inteligencia artificial son vinculantes, sobre las diferentes opciones disponibles y si tiene derecho a asesoramiento legal y derecho a acceder a un tribunal. También debe estar claramente informado de cualquier procesamiento previo de un caso por inteligencia artificial antes o durante un proceso judicial y tener el derecho de objetar, para que su caso pueda ser escuchado directamente por un tribunal".

Estamos a tiempo de aprovecharnos

No creo que la inteligencia artificial, de por si, constituya una seria amenaza para los derechos humanos. Es muy común afirmar cuando se señalan los peligros de la inteligencia artificial que ésta no va de tecnología sino de personas. En este sentido, la amenaza para los derechos puede venir del uso que las personas hagamos de la inteligencia artificial. De ese uso dependerá el que esta tecnología sea una oportunidad o un grave riesgo³⁷. De ahí la importancia de tomarnos en serio la tarea de establecer una regulación presidida por los derechos.

³⁷ En el momento en el que se entrega este texto para su publicación en los *Materiales de Filosofía del Derecho* estamos viviendo la pandemia del COVID-19, en la que de forma clara, la Inteligencia Artificial se presenta con esta doble cara de riesgo y oportunidad. Así, esta tecnología se utiliza para el intercambio

Ahora bien, esa tarea no tendrá éxito si no se toma en serio también la educación en derechos, esto es, si la educación digital no incorpora a los derechos humanos.

En España, es de celebrar que la Ley 3/2018 haya reformado la Ley de Educación y la Ley de Universidades. Ahora bien, aunque todavía es pronto para ver cómo se desarrolla y aplica la normativa, me parece que los esfuerzos principales en este campo se están dirigiendo hacia la adquisición de competencias y hacia el conocimiento de los riesgos personales.

No hay una mirada más general sobre riesgos sociales o un enfoque sobre el por qué y el para qué³⁸. Pero, para que la inteligencia artificial sea una herramienta de satisfacción de los derechos humanos, necesitamos una educación digital que tenga en cuenta los derechos humanos.

Y en este campo también contamos con algún texto que nos proporciona cierta esperanza. Se trata del Documento final de la Conferencia Internacional sobre la Inteligencia Artificial y la Educación, ‘‘Planificación de la educación en la era de la inteligencia artificial: dirigir los avances’’, celebrada entre el 16 y el 18 de mayo de 2019 en Pekín, y que lleva por título: ‘‘Consenso de Beijing sobre la inteligencia artificial y la educación’’³⁹.

En dicho documento, se reafirma ‘‘el enfoque humanista de la UNESCO respecto de la utilización de la inteligencia artificial con miras a proteger los derechos humanos y a proporcionar a todas las personas los valores y las competencias necesarios para una colaboración eficaz entre el ser humano y la máquina en la vida, el aprendizaje y el trabajo, y para el desarrollo sostenible’’. Y se afirma que ‘‘el desarrollo de la inteligencia artificial debe estar controlado por el ser humano y centrado en las personas; que

de conocimientos, para predecir la evolución de la pandemia, para apoyar a los sanitarios...; pero también se habla de los riesgos de su uso para la privacidad o en relación con el control de la población.

³⁸ Puede comprobarse esto en el informe *Digital Education at School in Europe. Eurydice Report*, European Commission/EACEA/Eurydice, Luxembourg, 2019. Disponible en: https://eacea.ec.europa.eu/national-policies/eurydice/sites/eurydice/files/en_digital_education_n.pdf (última consulta el 12 de diciembre de 2019).

³⁹ Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (última consulta el 12 de noviembre de 2019).

la implantación de la inteligencia artificial debe estar al servicio de las personas para mejorar las capacidades humanas; que la inteligencia artificial debe concebirse de manera ética, no discriminatoria, equitativa, transparente y verificable; y que el impacto de la inteligencia artificial en las personas y la sociedad debe ser objeto de seguimiento y evaluación a lo largo de las cadenas de valor”.