EDICIONES Universidad Valladolid

Recopilar y vigilar: algunas consideraciones filosófico-jurídicas sobre inteligencia artificial*

Collection and Surveillance: a legal-philosophical approach about Artificial Intelligence

ROGER CAMPIONE
Dpto. de Ciencias Jurídicas Básicas
Facultad de Derecho
Universidad de Oviedo
Dirección: Campus de El Cristo s/n, 33006 - Oviedo
Correo electrónico campione@uniovi.es
ORCID: 0000-0002-6331-5072

DOI: https://doi.org/10.24197/st.Extra_2.2021.123-139

RECIBIDO: 14/03/2021
ACEPTADO: 19/06/2021

Resumen: Las aplicaciones basadas en la inteligencia artificial abren la posibilidad concreta de replicar en máquinas una capacidad de razonamiento similar a la humana, otorgando a estas la facultad de construir representaciones no programadas gracias al aprendizaje proporcionado por algoritmos y Big Data. El abanico de inevitables dilemas que la aplicación de estos sistemas plantea en muchos ámbitos de la vida social resulta sobrecogedor y la dimensión normativa de la inteligencia artificial es evidentemente una de las más delicadas, por los intereses y principios que hay en juego. Se introducen aquí algunas consideraciones de orden iusfilosófico sobre las relaciones entre el derecho, la política y la vigilancia tecnológica.

Palabras clave: inteligencia artificial; privacidad; vigilancia; predicciones; sesgos.

Abstract: Applications based on artificial intelligence open up the concrete possibility of replicating in machines a reasoning ability similar to humans one, giving them the power to carry out non-programmed representations given that learning provided by algorithms and Big Data. The range of unavoidable dilemmas that the application of these systems raises in many areas of social life is overwhelming and the normative dimension of artificial intelligence is obviously one of the most delicate, due to the interests and principles involved. Some philosophical considerations on the relationships between law, politics and technological surveillance are raised here.

Keywords: Artificial Intelligence; privacy, surveillance; prediction; bias.

* Este trabajo se ha realizado en el marco del Proyecto "El logos de la guerra" (DER2017-82106-R), financiado por el Programa Estatal de I+D+i Orientada a los Retos de la Sociedad de la Agencia Estatal de Investigación.

1. DERECHO E INTELIGENCIA ARTIFICIAL: *POLICIES* DE UN MARCO PROBLEMÁTICO

We are all just prisoners here
Of our own device

The Eagles, *Hotel California*

Decía el genial matemático John von Neumann en los últimos años de su vida que el progreso tecnológico había sobrepasado la evolución moral y ética (Poundstone, 2005, p. 24). Eran los años cincuenta del siglo pasado. Hoy, en el primer tercio del siglo XXI, el desarrollo de la inteligencia artificial, la robótica y las nuevas tecnologías convergentes, avanza a un ritmo tan vertiginoso que afecta incluso a la evolución biológica del ser humano. Hay quien ha mantenido que los avances en inteligencia artificial, ingeniería genética y nanotecnología aumentarán de tal modo las capacidades físicas y cognitivas de los humanos que crearán una nueva especie generadora de un pensamiento biológico y a la vez no biológico (Kurzweil, 2005)[1].

Las aplicaciones basadas en la inteligencia artificial abren la posibilidad concreta de replicar en máquinas una capacidad de razonamiento similar a la humana, otorgando a estas la facultad de construir representaciones no programadas gracias al aprendizaje proporcionado por algoritmos y Big Data. Pese a que las investigaciones en el campo de la inteligencia artificial empezaron ya en la época de von Neumann y Turing, solo en los últimos años se ha producido un auténtico salto cualitativo en la posible consecución de resultados, cuando se ha abierto la perspectiva de aplicar a los sistemas de inteligencia artificial métodos de aprendizaje automático (*machine learning*) gracias a la capacidad de analizar ingentes masas de datos (*Big Data*). De este modo, la máquina no precisa que se le introduzca desde el exterior toda la información relativa a los datos, sino 'tan solo' el método para aprender a tratarla en función de la tarea que ha de llevar a cabo. Así, el sistema está capacitado para aprender conforme a pautas que pueden no quedar claras ni siquiera para su constructor humano.

El abanico de inevitables dilemas que la aplicación de estos sistemas plantea en muchos ámbitos de la vida social resulta sobrecogedor y la dimensión normativa de la inteligencia artificial es evidentemente una de las más delicadas, por los intereses y principios que hay en juego. A lo largo de su historia, las relaciones entre inteligencia artificial y derecho han pivotado en torno a dos ejes intersecados: por un lado, la elaboración de modelos computables de conocimiento jurídico, formalizando contenidos normativos a los que aplicar métodos de inferencia automática basándose en reglas, estudio de casos y doctrinas; por el otro, la reflexión sobre las implicaciones éticas ligadas al impacto jurídico de la inteligencia artificial, al fin de adaptar las

[1]Kurzweil sitúa en 2029 la simulación funcional de la inteligencia humana (Kurzweil, 2005: p. 199).

normas y las instituciones a ciertos cambios que inevitablemente afectan a valores y principios que los ordenamientos procuran preservar (Sartor, 2020). El presente artículo se enmarca en este segundo aspecto.

Como prueba de la urgencia de regulación que afecta a las aplicaciones de la inteligencia artificial, recientemente se han intensificado las *policies* comunitarias sobre la materia. Consignando una definición del concepto de 'inteligencia artificial', la Comisión Europea ha manifestado que "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– con el fin de alcanzar objetivos específicos". Estos sistemas, una "vez que funcionan bien, pueden ayudar a mejorar y automatizar la adopción de decisiones en el mismo ámbito. Por ejemplo, un sistema de IA se puede entrenar con vistas a utilizarlo para detectar los ataques informáticos a partir de los datos obtenidos de la red o del sistema en cuestión". En la misma sede, también ejemplifica el funcionamiento de tales sistemas: "[l]os 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). Estamos utilizando la IA diariamente, por ejemplo, para traducir de un idioma a otro, generar subtítulos en los vídeos o bloquear el correo electrónico no solicitado (spam)" (Comisión Europea, 2018).

Más recientemente, el *High-Level Expert Group on Artificial Intelligence*, nombrado por la Comisión Europea, ha redactado unas "Directrices éticas para una Inteligencia Artificial fiable", que plantean una serie de exigencias consideradas insoslayables, para compaginar las potencialidades de las nuevas tecnologías con la necesidad de regular sus aplicaciones sin que peligren los cimientos de nuestros sistemas normativos. Resumidamente, los llamamientos persiguen satisfacer: a) la necesidad de la acción y el control humano para evaluar los efectos de los sistemas de inteligencia artificial sobre los derechos fundamentales, controlando la asignación de tareas entre sistemas de inteligencia artificial y trabajadores humanos y midiendo el nivel adecuado de control humano en los casos específicos de aplicación de sistemas de inteligencia artificial; b) el análisis de la vulnerabilidad de estos sistemas para garantizar su solidez técnica y su seguridad, adoptando medidas para evitar usos no deseados, asegurando planes de contingencia y definiendo protocolos de precisión, fiabilidad y reproducibilidad; c) la protección de la privacidad en el manejo de los datos, la búsqueda de su calidad, integridad y gobernanza (quiénes, cómo, cuándo y con qué propósito pueden acceder); d) la transparencia de los sistemas de inteligencia artificial, garantizando su trazabilidad (los métodos empleados para su diseño, desarrollo y validación de resultados), su interpretación (tratando de utilizar el modelo más sencillo de explicación comprensible) y el establecimiento de mecanismos de información para los usuarios sobre las razones, los criterios y los beneficios de los productos o los servicios, advirtiendo con claridad de las posibles

carencias de los sistemas; e) la necesidad de evitar sesgos o discriminaciones injustas, teniendo en cuenta la diversidad de los usuarios, el impacto que pueden tener en ellos los sistemas de inteligencia artificial y la garantía de la accesibilidad universal para la información y la utilización de sus resultados; f) la necesidad de medir y reducir el impacto ambiental de los sistemas de inteligencia artificial, de asegurar la correcta comprensión de sus efectos sociales, en el caso que interactúen directamente con seres humanos, y la evaluación de su impacto social global en términos democráticos; g) el establecimiento de mecanismos de rendición de cuentas en cuanto a la auditabilidad de estos sistemas, la minimización y notificación de sus efectos negativos, la documentación del equilibrio entre intereses y valores implicados y, a este respecto, la instauración de mecanismos que permitan obtener compensaciones (Grupo de expertos de alto nivel sobre inteligencia artificial, 2019, pp. 33-41).

En el plano nacional, la *Carta de derechos digitales*, expuesta por el Gobierno español en su primera versión para la consulta pública el 17 de noviembre de 2020, insiste en la implementación de directrices que garanticen la protección de los derechos individuales en el entorno digital. El apartado XXIII, en concreto, se refiere a los "Derechos ante la Inteligencia artificial" con el siguiente tenor:

"1. En el desarrollo y ciclo de vida de los sistemas de Inteligencia Artificial: a) Se deberá garantizar el derecho a la no discriminación algorítmica, cualquiera que fuera su origen, causa o naturaleza del sesgo, en relación con las decisiones y procesos basados en algoritmos. b) Se asegurarán la transparencia, auditabilidad, explicabilidad y trazabilidad. c) Deberán garantizarse la accesibilidad, usabilidad y fiabilidad.

2. Las personas tienen derecho a no ser objeto de una decisión basada únicamente en procesos de decisión automatizada, incluidas aquéllas que empleen procedimientos de inteligencia artificial, que produzcan efectos jurídicos o les afecten significativamente de modo similar, salvo en los supuestos previstos en las leyes. En tales casos se reconocen los derechos a: a) Solicitar una supervisión e intervención humana; b) Impugnar las decisiones automatizadas o algorítmicas.

3. Se deberá informar a las personas sobre el uso de sistemas de Inteligencia Artificial que se comuniquen con seres humanos utilizando el lenguaje natural en todas sus formas. Deberá garantizarse en todo caso la asistencia por un ser humano a solicitud de la persona interesada.

4. Se prohíbe el uso de sistemas de Inteligencia Artificial dirigidos a manipular o perturbar la voluntad de las personas, en cualesquiera aspectos que afecten a los derechos fundamentales"[2].

De todos modos, conviene una advertencia: por muy sofisticados que sean los sistemas y dispositivos creados hasta ahora, sus capacidades son parciales. No manejan el sentido común y las variables contextuales. De hecho, el asombroso

[2]El "Documento para la consulta pública. Carta de los derechos digitales" puede leerse en https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/participacion_publica/audiencia/ficheros/SEDIACartaDerechosDigitales.pdf [fecha de consulta: 18 de noviembre de 2020].

avance que ha habido durante las últimas décadas en ingeniería informática, no se ha visto acompañado por un desarrollo mínimamente parecido de niveles de conciencia informática. Los ordenadores de la época de Turing no son en este sentido muy distintos de los actuales. Estaríamos ante una reedición *millenial* del mito de Prometeo relatado por Platón en el *Protágoras*: las máquinas que poseen el mayor conocimiento técnico imaginable, compuesto por datos y números, carecen del más elemental conocimiento moral, el sentido del pudor y la justicia, base esencial para tomar decisiones que se puedan considerar 'buenas' desde el punto de vista moral y político.

Sin embargo, como recuerda Harari, la novedad importante inducida por los avances cosechados por las aplicaciones de la inteligencia artificial consiste en el peligro de que los seres humanos puedan perder valor "porque la inteligencia se está desconectando de la conciencia" (Harari, 2016, p. 341).Teniendo en cuenta esto, si bien resulta (al menos por ahora) ciencia ficción la posibilidad de construir una máquina a la que atribuir responsabilidades de tipo consecuencial por sus acciones, sin que concurra en tales cadenas de actos un ser humano, ya que para ello se requeriría una fiabilidad contrastada y cierto grado de empatía que le permitiera formular un juicio sobre las acciones ajenas, hace ya tiempo que, como se acaba de esbozar, ha sonado la alarma ética y jurídica en las altas esferas nacionales y comunitarias.

Hay ámbitos muy sensibles que implican riesgos a tener en cuenta al lado de las indudables potencialidades ofrecidas por los nuevos sistemas de inteligencia artificial. Los ámbitos especialmente vulnerables, en este sentido, serían la privacidad, puesta a prueba por la cantidad de información personal que cabe inferir de nuestras actividades más 'inocuas' en la red; la opacidad, es decir, el hecho de desconocer cómo funcionan los dispositivos y los programas con los que interactuamos; los sesgos y las potenciales discriminaciones provocadas por algoritmos diseñados por seres humanos; la asimetría entre sujetos individuales y entre sujetos colectivos que siempre refleja la distinta capacidad de acceso a los recursos, particularmente incisiva en el caso del desarrollo tecnológico; los aspectos éticos involucrados por las aplicaciones de la inteligencia artificial, no necesariamente resueltos por su regulación legal; y la veracidad, puesta en tela de juicio gracias a la generación de una realidad virtual que con las posibilidades de la inteligencia artificial se torna indistinguible de la realidad actual (Fernández-Galindo, 2019, p. 69).

El eje 'privacidad de la información/predictividad de las conductas' me parece un vector adecuado para tocar de manera transversal muchos de los aspectos involucrados.

2. LA OBSOLESCENCIA DE LA PRIVACIDAD Y EL GRAN RECOPILADOR

El ritmo vertiginoso al que avanza el desarrollo de las tecnologías convergentes tiene efectos potencialmente disruptivos sobre la eficacia de las normas vigentes. La privacidad es un ejemplo paradigmático de la desconexión entre validez y eficacia de la ley, causada por la creciente aceleración de las aplicaciones ligadas a la inteligencia artificial. La velocidad a la que avanzan las tecnologías de la comunicación pone en fuera de juego las legislaciones que se van elaborando para disciplinar la nueva realidad digital.

Para intentar tutelar la privacidad, en este contexto de manejo masivo de información personal ha entrado en vigor el Reglamento General de Protección de Datos de la Unión Europea 2016/679 (RGPD), que ha endurecido los requisitos para que el consentimiento del interesado permita el tratamiento de sus datos personales. El apartado 11 del art. 4 define el consentimiento del interesado como "toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen". Pero, como es habitual en derecho, se trata de una previsión normativa que crea una ficción de consentimiento informado, porque lo que aceptamos libremente con un clic o como *default* de cookies es, para el agente medio, sustancialmente un misterio. El mecanismo funciona como el del art. 6.1 del Código Civil: así como conocemos todas las normas jurídicas vigentes porque el ordenamiento jurídico *presume* tal cosa, estableciendo que la ignorancia no excusa del cumplimiento, al aceptar libre y voluntariamente con un clic un tratamiento determinado de nuestros datos significa, por arte de birlibirloque, que *sabemos exactamente* qué ocurre con nuestra información, cómo viene almacenada y qué usos la empresa o los sujetos en cuestión harán de ella. Naturalmente, no es cierto.

De ahí que los esfuerzos profusos desde los años noventa del siglo pasado para elaborar reglas que permitieran disciplinar las nuevas realidades digitales se hayan visto abocados a un sustancial fracaso, en el sentido de que el derecho no ha logrado garantizar el objetivo, ni la ciencia jurídica ha sabido captar cabalmente el agrietamiento del concepto mismo de *privacidad* subyacente en la cultura jurídica de los ordenamientos liberales (Romeo, 2020, p. 108). Toda la información que generamos en el espacio digital, desde el correo electrónico al whatsapp, desde las transacciones virtuales a las redes sociales, adquiere valor porque existen herramientas analíticas que hacen posible 'ubicarla' sin límite. Y no se trata de un bien valioso que cada uno almacena debajo de su colchón, sino de una especie de moneda encriptada que puede alimentar dos grandes 'negocios' biopolíticos de titularidad ajena: uno público, la seguridad, y otro privado, el comercio de macrodatos personales. Con los algoritmos inteligentes es posible procesar, seleccionar, clasificar todo tipo de información sin perder una mínima parte. Una suerte de anamnesis platónica invertida, donde el aprendizaje ya no consiste en recordar porque nada es olvidado. La inutilidad de la reminiscencia convierte la personalidad humana en absolutamente transparente, pues nada se le puede ocultar al

Gran Recopilador. De hecho, el *Big Data* implica de por sí la negación de la privacidad y la conversión de la persona de partícula individual a flujo de informaciones. Como si el ser humano transitara a otro estado cuántico, el ondulatorio e inmaterial: pura energía al servicio del Gran Recopilador.

La dimensión biopolítica de este proceso se está notando especialmente en estos tiempos de pandemia marcados por una lógica de la emergencia que ha convertido en habitual un estado de excepción que intentamos exorcizar llamándolo 'nueva normalidad'. La pérdida de privacidad implica el crecimiento del control y me parece que toca una fibra sensible Byung-Chul Han cuando apunta que con la pandemia nos dirigimos hacia un régimen de vigilancia biopolítica cuyo objeto no son solo las comunicaciones, sino nuestro cuerpo, nuestra salud, en definitiva, nuestros indicadores biométricos. Aunque el control masivo mediante el uso de la tecnología no es obviamente algo nuevo, pues ya sabemos que también las grandes multinacionales nos vigilan como consumidores para influir en nuestros gustos y preferencias, también Harari destaca este elemento sobrevenido. Hasta hoy las nuevas tecnologías se empleaban para saber qué páginas web visitamos o qué productos buscamos en la red. Sin embargo, a raíz del coronavirus, es posible que cuando pinchemos en un link lo que se vaya a detectar no sea solo la dirección a la que estamos accediendo, sino la temperatura del dedo que ha tocado la pantalla, la presión sanguínea, el latido cardíaco y todos los datos biométricos del caso (Harari, 2020). Todo esto es posible gracias a que ciertos algoritmos pueden analizar un volumen inimaginable de información y podrían descubrir una enfermedad en un sujeto antes de que él mismo se dé cuenta o advierta algún síntoma. Una eficacia total en salud pública, a cambio de la aniquilación de la privacidad. Nos hallamos ante una de las grandes disyuntivas político-jurídicas que van a marcar el mundo post-pandemia y el problema surge por el increíble avance de las tecnologías ligadas a la inteligencia artificial y el Big Data, sin los cuales sería impensable adentrarse tanto en la esfera jurídica personal. Está en lo cierto quien avisa de que "[p]lantear este tipo de debates en términos binarios es peligroso e incluso demagógico" (Cotino Hueso, 2020, p. 3), pero también creo que la propia lógica del estado de excepción utilizada como dispositivo para afrontar la epidemia (en un sentido puramente schmittiano, como suspensión constitucionalmente prevista del ordenamiento jurídico) inevitablemente 'radicaliza' los polos de la alternativa y nos acerca un poco más a la visión de Baudrillard de un universo no dialéctico, sino condenado a los extremos, no al equilibrio; destinado al antagonismo, no a la conciliación ni a la síntesis (Baudrillard, 1984, p. 5).

Tal vez hoy una interpretación realista de los acontecimientos planetarios sea la de Han, quien sostiene que la estrategia ganadora contra el Covid-19 parecen tenerla los países asiáticos, precisamente porque su cultura menos individualista y más autoritaria les ha permitido adoptar medidas eficaces respecto del contagio, gracias a una actitud para nada axiomática frente a las libertades jurídicas subjetivas (Han, 2020). Sin embargo, no se puede negar que en Europa y en el mundo occidental se

han tomado medidas sin precedentes que quiebran el molde del Estado de derecho y se suman a una metamorfosis regresiva de la ciudadanía democrática, fraguada en el ciberespacio de la "red social"[3]. El tránsito a un nuevo estatuto de la excepcionalidad, después de un par de décadas rodeados de retórica internacional de emergencias construidas para alimentar la llamada "guerra contra el terror", agudiza la preocupación. Nada ha vuelto a ser igual en los aeropuertos después del 11-S… y puede que nada vuelva a ser igual a la época del Estado de derecho, una vez que se haya superado la pandemia, más aún considerando la adaptación generalizada a las medidas hodiernas, vividas por la población de los países de tradición liberal y democrática como el precio a pagar para vivir en seguridad.

Son metamorfosis del concepto de seguridad: con el paso de la biopolítica a la *psicopolítica*, descrito por Han, se manifiesta el *modus operandi* de un sistema neoliberal que en el siglo XXI poco tiene que ver con el capitalismo productivo. El objetivo del control social alcanzable gracias a las nuevas tecnologías ya no es solo biopolítico, el control del cuerpo, sino psicopolítico, la programación de la mente, el manejo de las emociones (Han, 2014). No sería viable este escenario de manipulación personal profunda sin la intervención de algoritmos inteligentes y el *Big Data*[4]. Sin la inteligencia artificial, el Gran Recopilador quizá no estaría desnudo, pero sí en pañales. Probablemente, solo la *STASI*, con 90.000 empleados y una red de 170.000 informadores sobre una población de 17 millones de habitantes, es decir, con un espía cada 65 ciudadanos, podría representar dignamente al antecesor de la vigilancia digital (Tamburrini, 2020). Estos sistemas de control y monitorización mediante el análisis de datos y comportamientos individuales, por tanto, se utilizan también en las democracias occidentales con, entre otros, el fin de influenciar secretamente las opiniones y las manifestaciones de voto individuales, como demuestra el caso Facebook/Cambridge Analytica, erosionando así la autonomía y los derechos políticos de los ciudadanos (Zuboff, 2020b, pp. 345-393). Ha tenido cierta repercusión la revelación de una propuesta en el seno de la Unión Europea para que las indicadas genéricamente como "autoridades competentes" (no parece descabellado pensar en servicios de inteligencia y espionaje) puedan tener acceso a las comunicaciones encriptadas por el sistema *end to end* que los usuarios solemos utilizar en los mensajes que intercambiamos a diario en *apps* como Whatsapp. El documento, publicado por el portal austriaco RadioFM4, apunta a la necesidad de revisar los marcos legales en la materia al fin de que las imprecisadas autoridades competentes tengan acceso al contenido de los mensajes encriptados[5]. Es decir, se solicita un acuerdo para la vigilancia masiva de las conversaciones privadas, en contra de los principios impulsados por el RGPD, conforme a los cuales el cifrado es

[3] Sobre la degradación de la ciudadanía en el marco de la revolución digital, *vid.* Pietropaoli, 2021.
[4] Para una breve referencia a las posturas críticas sobre lo que podríamos llamar "el imperio de los datos", *vid.* Cotino Hueso, 2019, pp. 13-15.
[5] El artículo que hace pública la propuesta puede leerse en la web de RadioFM4 (https://fm4.orf.at/stories/3008930/), mientras que el documento en cuestión está disponible en https://files.orf.at/vietnam2/files/fm4/202045/783284_fh_st12143-re01en20_783284.pdf

un elemento básico para garantizar la privacidad de las comunicaciones personales. Paradójicamente (o no tanto…), a medida que estas garantías se difuminan al ser consideradas un obstáculo al orden público por parte de los gobiernos democráticos, aumentan y se fomentan prácticas dirigidas a un empoderamiento de la intimidad según criterios que, en mi opinión, malinterpretan el valor jurídico de la privacidad[6].

3. PREDICCIONES Y SESGOS

A medida que se incrementa la utilización de sistemas artificiales autónomos puede que se reduzcan los niveles de autonomía humana. Esta pérdida de independencia del ser biológico con respecto al silícico es debida, entre otros factores, como se ha visto, a la extraordinaria capacidad de recopilación de la información gracias a algoritmos que permiten clasificar a las personas por sus preferencias, gustos y rutinas de navegación y no por acciones concretamente realizadas en el mundo físico. Es lo que se conoce como *googlelización* de la identidad, aupada por un nuevo ambiente global en el que nos hallamos inmersos, la *infoesfera*, en la que perdemos nuestra condición de ente individual delimitado para convertirnos en *inforg*, un flujo interconectado de informaciones, un ser mestizo constituido por agentes biológicos y artefactos ingenieriles (Floridi, 2012, p. 11).

Este nuevo *nomos* no indica una pura subdivisión territorial, sino un ciberespacio en el que la personalidad humana se compone de números procedentes de un flujo de información. Así es como se consuma el paso de la revolución industrial a la revolución digital. No todo es negativo, obviamente: limitándonos a un ámbito del derecho, bastaría con mencionar las ventajas de los programas dotados de inteligencia artificial para el análisis y sistematización del enorme volumen de información relevante para la actividad jurídica y la práctica profesional del derecho (Solar Cayón, 2019). Sin embargo, gracias al desarrollo del internet de las cosas los datos relativos a nuestra vida cotidiana, a nuestros movimientos tanto físicos como en la red, quedan recogidos de forma digital y son susceptibles de uso comercial mediante la aplicación de ciertos algoritmos.

Zuboff ya había hablado del "capitalismo de la vigilancia", una nueva era de la economía política caracterizada por una lógica emergente de acumulación en la esfera digital, donde el flujo de datos personales recopilados, más allá de su uso para mejorar productos y servicios, cobra valor mercantil por su notable funcionalidad predictiva. Gracias a los sistemas de inteligencia artificial estos datos se convierten en objeto de comercialización en lo que esta autora define como "mercado de futuros conductuales", cuya primera mercancía fueron los clics de Google. Este capitalismo

[6] Esta consideración no es más que una percepción personal derivada de mi profesión, la docencia universitaria, en cuyo ámbito se percibe con cada vez más frecuencia la obligación o la recomendación de cumplir exigencias legales de anonimización, incluso en materias como las calificaciones de los estudiantes, en las que la publicidad de los datos identificativos debería constituir una de las mayores garantías de control colectivo de que la actividad docente se ajusta a los principios normativos de la profesión y a los preceptos constitucionales. Por encima de todos, a la igualdad de trato y a la prohibición de discriminaciones injustificadas.

de la vigilancia basa su estructura en opacos mecanismos de extracción de información, mercantilización y control que embotan los principios normativos democráticos empleando nuevas asimetrías de conocimiento y poder que amplían las desigualdades, favorecidas por la ilegibilidad de los procesos automatizados. Vuelve así la pérdida de la privacidad, producida por una dependencia del mundo digital que nos rastrea y manipula de forma habitual y nos conduce a un estado de resignación en el que renunciamos a la privacidad, bajo nuestra convicción de no tener nada que ocultar, sin reparar en que nos están imponiendo una elección ilegítima (Zuboff, 2015; 2020a)[7].

Las empresas tecnológicas que registran nuestra actividad personal en la web utilizan esa información para dirigir la atención del usuario hacia determinadas noticias y anuncios. Así, en las redes sociales, la publicidad es distribuida a través de algoritmos que la envían de forma segmentada a partir de nuestros patrones de navegación. Y la inteligencia artificial también está detrás de los procedimientos que muchas empresas utilizan para seleccionar a sus trabajadores o de la predicción de las hipotecas. Y, más allá del ámbito comercial, se aprecia la importancia de la información biométrica en aplicaciones como la optimización de las rutas que realizamos, el reconocimiento facial o los diagnósticos a partir de los datos de los pacientes. Los peligros derivados de un uso sesgado de tal volumen de información son evidentes y por eso la Agencia Española de Protección de Datos (AEPD) ha elaborado un documento guía para la adecuación al RGPD de tratamientos que incorporan inteligencia artificial, con el objetivo de aclarar las dudas que esta genera en usuarios, especialistas e instituciones, con respecto a los derechos involucrados y a las necesidades de seguridad jurídica de todos los intervinientes en los procesos que emplean tratamiento de datos mediante técnicas basadas en el aprendizaje automático[8].

En muchos países están de moda programas predictivos de delitos mediante algoritmos capaces de procesar datos históricos individuales para diseñar mapas de delincuencia en distintas ciudades. En algunos de estos modelos, para calcular los índices, se pregunta por los antecedentes penales de amigos y familiares o por la primera vez que se tuvo trato con la policía y las respuestas, como es de esperar, varían sobre la base de circunstancias sociales, es decir, del nivel de vida o de marginación social de unos barrios frente a otros. El LSI-R (*Level of Service Inventory-Revised*), por ejemplo, es un modelo matemático de peligrosidad y reincidencia basado en cuestionarios de preguntas. Si bien estas representaciones no

[7] En palabras de Zuboff, este capitalismo de la vigilancia "[n]os predispone a racionalizar la situación con resignado cinismo y a crear excusas que funcionan como mecanismos de defensa («tampoco tengo nada que ocultar»), cuando no hallamos otras formas de esconder la cabeza y optar por la ignorancia para afrontar la frustración y la impotencia. Por esa vía, el capitalismo de la vigilancia nos impone una decisión fundamentalmente ilegítima que los individuos del siglo xxi no deberíamos tener que tomar, y cuya normalización hace que, finalmente, no solo estemos encadenados, sino que también vivamos contentos de estarlo" (Zuboff, 2020b, p. 25).

[8] El documento, "Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción", está disponible en https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

pueden contener referencias a criterios como la raza porque serían ilegales, pueden incurrir igualmente en retroalimentaciones sesgadas. Un sujeto procedente de un barrio de 'alto riesgo' de reincidencia, donde es más probable que algún conocido haya tenido roces con la ley, se verá expuesto a condenas más largas y si pasa más tiempo en la cárcel, más difícil se tornará su reinserción en la vida laboral tras la salida. Con ello, también aumentarán las probabilidades objetivas de que vuelva a delinquir. Si nos atenemos a un estricto criterio matemático, resulta que el programa supo prever el resultado, pero, de hecho, el propio modelo utilizado habrá desempeñado un papel activo en causarlo (O'Neil, 2018, pp. 36-38).

El problema, además, no reside tan solo en los posibles sesgos racistas o clasistas implícitos en el diseño de estos sistemas inteligentes. Existen dudas también acerca de su capacidad efectiva para predecir los comportamientos y las tendencias; así, se ha mostrado que modelos como el COMPAS, un programa predictivo muy usado en los Estados Unidos, tampoco acierta en sus previsiones, incluso si se lo compara con personas dotadas de conocimientos escasos en la materia (Dressel y Farid, 2018)[9]. Hay una inversión lógica en un patrón semejante, puesto que se estima probable que alguien cometa un delito porque un determinado indicador basado en un algoritmo lo califica como peligroso; sin embargo, la arquitectura jurídica del Estado de derecho se sustenta en el principio garantista opuesto, para el cual en tema de reincidencia la peligrosidad suele ser la consecuencia y no la causa del delito. Estos modelos parecen así más propios de la sociedad representada en *Minority Report,* donde la culpabilidad no depende de hechos sino de predicciones, en nuestro caso, algorítmicas[10]. Como se ha destacado, el primer paso para juzgar a las personas no por lo que han hecho sino por lo que potencialmente podrían hacer, consiste en servirse de herramientas de inteligencia artificial como el COMPAS en el ámbito de las medidas cautelares (Nieva Fenoll, 2018, p. 74). Con el agravante de que se trata de instrumentos de predicción opacos porque sus creadores, amparándose en la propiedad intelectual no ofrecen la información necesaria para comprender cabalmente su funcionamiento (admitiendo que eso fuera completamente posible). Es conocido el caso de un ciudadano americano, Eric Loomis, que fue condenado en 2013 por no haberse parado ante la orden de un policía mientras conducía un coche sin el permiso del propietario. Loomis fue condenado a seis años de cárcel y cinco de libertad vigilada por la Corte del Condado de La Crosse, decisión luego confirmada por la Corte Suprema de Wisconsin. El acusado, se lee en la sentencia, había sido considerado como sujeto con alta probabilidad de reincidencia por el programa COMPAS. Y cuando Loomis solicitó el acceso al código fuente del software para conocer las razones de la decisión algorítmica y poder contra-argumentar, se le denegó porque el algoritmo estaba cubierto por la propiedad intelectual al pertenecer

[9] Nieva Fenoll recuerda que los índices de peligrosidad calculados por un grupo de voluntarios demostraron una eficacia del 67%, superior en dos puntos a la de COMPAS (Nieva Fenoll, 2018, p. 70 y la bibliografía allí reseñada).

[10] En otro lugar me he ocupado del tránsito del Estado social de derecho, basado en el principio liberal del daño, al Estado preventivo, caracterizado por el principio post-liberal del miedo (Campione, 2020, pp. 27-35).

a una empresa privada (Equivant, antes Northpointe) (*State v. Loomis*, 881 N. W.2d 749 (Wis. 2016))[11]

Rodotà ya había hablado de la expropiación de la política por parte de la tecnociencia (Rodotà, 2018, p. 138) y entra en juego algo más, porque la cuestión decisiva no tiene que ver con cómo las tecnologías convergentes afectan a nuestra relación con nosotros mismos, sino a cómo modifican la relación que tenemos con el poder (Amato, 2014, p. 188). De ahí la pregunta relativa a si el acceso, la aplicación y las responsabilidades de las nuevas tecnologías convergentes serán codificadas a través de la lógica de los derechos o de los mercados. Cobra sentido la sospecha de que la semántica normativa (también) de la inteligencia artificial dependa principalmente del paradigma Humpty Dumpty: más que averiguar las distintas cosas que se pueden hacer, lo que importa en primer lugar es saber quién manda.

Las distorsiones que retroalimentan los sesgos predictivos provocados por este tipo de sistemas que emplean la inteligencia artificial en el análisis del *Big Data*, pueden verificarse incluso cuando el criterio de clasificación de datos no se basa en el perfil de las personas, sino en la ubicación geográfica, para así delinear mapas predictivos de delitos. En programas como el PredPol, usado en Estados Unidos, el problema estriba en la selección de los delitos indicados en la configuración inicial: si, por ejemplo, en la categoría de delitos violentos se incluyen los leves de alteración del orden público, la geografía predictiva de criminalidad varía notablemente. Estos delitos de menor intensidad son ciertamente más frecuentes en zonas o barrios determinados, a diferencia, por ejemplo, de los delitos de cuello blanco, relacionados con la actividad económica y financiera. Este tipo de operaciones se desarrollan en áreas urbanas diferentes y es bastante más probable que allí se cometan tales delitos. Los mapas de criminalidad, por tanto, cambian notablemente según los inputs iniciales, esto es, dependiendo de los marcadores relevantes introducidos en el sistema por los gestores (humanos) del programa. Porque el mapa diseñado con la inclusión de los delitos leves de alteración del orden público, conducirá a un más intenso control policial en ciertas zonas. A medida que haya más patrullas vigilando esos barrios, aumentará inevitablemente el número de infracciones penales registradas y, por ende, el resultado del modelo predictivo no será una consecuencia neutral de la aplicación de un algoritmo matemático, sino el producto de sus propios sesgos de partida (O'Neil, 2018, pp. 107-115).

Los programas predictivos han incurrido en violaciones de derechos incluso cuando se han aplicado a otro tipo de delitos. En los Países Bajos, el Tribunal de Distrito de La Haya ha considerado que SyRI (acrónimo de *System Risk Indication*), un sistema de análisis utilizado por el Ministerio de Asuntos Sociales y Empleo para detectar fraudes al Estado, viola el derecho al respeto de la vida privada, reconocido en el art. 8 del Convenio Europeo de Derechos Humanos. Los algoritmos utilizados por SyRI calculan quién tiene más probabilidades de defraudar al Estado analizando datos económicos de los contribuyentes como los impuestos, las multas, los seguros

[11] Sobre este caso y el impacto jurídico del COMPAS *vid.* Simoncini y Suweis, 2019, pp. 94 ss.

y otro tipo de información. Según el Tribunal, en el juicio de ponderación entre la persecución del interés público asociado al uso de estas tecnologías –prevenir y combatir el fraude– y la interferencia que se produce con el respeto a la vida privada, prevalece en este caso la tutela del derecho a la privacidad reconocido en el art. 8 párrafo 2 del CEDH, en el sentido de que el instrumento legal empleado para evitar delitos puede presentar sesgos de estigmatización y discriminación de la ciudadanía, debido a la masa de información individual y colectiva que maneja[12]. Este caso ha atraído la atención del *special rapporteur* de la ONU sobre pobreza extrema y derechos humanos, Philip Alston,que ha actuado como *amicus curiae* del Tribunal neerlandés. Alston ha destacado que *SyRI* ha sido utilizado predominantemente en áreas urbanas donde es más alta la concentración de grupos vulnerables y con escasos ingresos, incluyendo vecindarios 'pobres' de ciudades como Capelle aan den Ijssel, Eindhoven, Haarlem y Rotterdam. Y en este caso, la opacidad del sistema y la falta de trasparencia en la accesibilidad a la información relativa al funcionamiento de los algoritmos no se debe a la voluntad de tutelar la propiedad intelectual, sino a la intención del Gobierno de no proporcionar información sobre cómo funcionan estos sistemas, para impedir que los potenciales violadores (que, señala Alston, en la norma que regula la aplicación de SyRI también permanecen *unclear*) dispongan de elementos que le faciliten formas de 'ganar al sistema' (Alston, 2019).

4. CONCLUSIÓN

Ya solo basándonos en estas pinceladas sobre ámbitos concretos, parece natural la preocupación que suscitan las aplicaciones de inteligencia artificial al campo jurídico, por la posible vulneración de principios y paradigmas normativos tradicionalmente propios de los ordenamientos democráticos y liberales. Las normas adoptadas para regular estos fenómenos, pensemos en la disciplina de la privacidad procedente del RGPD, indudablemente tienen como objetivo diseñar un marco de protección para ciertos derechos fundamentales. Sin embargo, no es tarea fácil reglamentar las condiciones y los efectos de decisiones tomadas en virtud de algoritmos cuyo funcionamiento ni tan siquiera es conocido del todo por sus programadores y que, además, pueden cambiar gracias a la capacidad del propio sistema de aprender por sí mismo (*deep learning*) (Gascón Marcén, 2020, p. 347).

La irrupción de la pandemia causada por el Covid-19 ha agudizado ciertas dudas ligadas a los límites del control social. La tecnología de la vigilancia ha descubierto un laboratorio inmejorable con la implementación de *apps* de geolocalización y *contact tracing* para la recopilación de datos y la extracción de información para hacer frente a las consecuencias del coronavirus. Si bien la utilización por parte de los usuarios ha estado en muchos casos, como el de la *app* española Radar COVID, muy por debajo de los umbrales de eficacia y de las expectativas (tampoco se ha

[12] Sentencia de 5 de febrero de 2020 de la Corte de Distrito de La Haya (Rechtbank Den Haag), Referencia: ECLI:NL:RBDHA:2020:1878 (versión inglesa que se puede consultar aquí).

impulsado una gran campaña de concienciación por parte de las autoridades), no se pueden ignorar las potencialidades de estas técnicas. De hecho, para poder llevar a cabo a gran escala esta actividad de vigilancia, han sellado una alianza empresas como Google y Apple, tradicionalmente enfrentadas en el mercado, comprometiéndose a proporcionar de forma gratuita los servicios y las funcionalidades de rastreo. Se asegura que se trata de una tecnología *opt-in*, es decir, los usuarios tienen que elegirla de manera explícita, y garantiza la privacidad y la transparencia, pero es natural que tales dispositivos de vigilancia representen y sean percibidos como una amenaza para los derechos de las personas. No hay duda de que deben buscarse soluciones jurídicas que permitan maximizar la eficacia contra el coronavirus y minimizar el impacto en la libertad y privacidad (Cotino Hueso, 2020, p. 3). Al mismo tiempo, sin embargo, hay que vigilar a los vigilantes que emplean las nuevas tecnologías para consolidar bienes colectivos como la salud y la seguridad. No solo porque, como se decía, no se sabe cómo serán aplicados técnicamente ni se conoce el nivel de transparencia de sus códigos o algoritmos, sino por el desplazamiento de poder social que implica semejante gestión de las identidades digitales por parte de actores privados que ocupan una posición destacada en los mercados. Al menos, si queremos seguir presumiendo de ser hijos del *siècle des Lumières*, del *Rule of Law* y, por decirlo con Bobbio, de la *Età dei diritti*. De lo contrario, podría ser que la ideología liberal de los derechos se sustentara, tal como sostiene Žižek mencionando la institución amish del *rumspringa*, en una pseudoelección, es decir, una elección formalmente libre en la cual, sin embargo, "las condiciones en las que deben tomar su decisión hacen que ésta no sea libre" (Žižek, 2005, p. 118). Pero, de ser así, nos lo deberíamos decir bien claro ante el espejo.

REFERENCIAS BIBLIOGRÁFICAS.

Alston, Ph. (2019), *Brief by the UNITED NATIONS SPECIAL RAPPORTEUR on extreme poverty and human rights as* Amicus Curiae *in the caseof NJCM c.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/ HA ZA 18/388)*, pp. 1-12 (https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf)

Amato, S. (2014), Neuroscienze e utilizzazione militare delle tecniche di potenziamento umano. *Etica & Politica*, XVI, 2, pp. 182-198.

Baudrillard, J. (1984), *Las estrategias fatales*, Barcelona: Anagrama; ed. or. *Les estatégies fatales*, París: Grasset, 1983.

Campione, R. (2020), *La plausibilidad del derecho en la era de la inteligencia artificial. Filosofía carbónica y filosofía silícica del derecho*, Madrid: Dykinson.

Comisión Europea (2018), Inteligencia artificial para Europa. *Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones*, COM(2018) 237 final.

Cotino Hueso, L. (2019), Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. Enfoques, modelos y principios de la respuesta del derecho. *Revista general de derecho administrativo*, 50, 1-37.

Cotino Hueso, L. (2020), Inteligencia artificial, *big data* y aplicaciones contra la COVID-19: privacidad y protección de datos. *Revista d'internet, dret i política*, 31, 1-17.

Dressel, J., Farid, H. (2018), The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4, n. 1, DOI: 10.1126/sciadv.aao5580: http://advances.sciencemag.org/content/4/1/eaao5580/tab-pdf

Fernández-Galindo, J.C. (2019), "Entrevista: Nuria Oliver". *Muy Interesante*, núm. 462, noviembre de 2019, pp. 66-71.

Floridi, L. (2012), *La rivoluzione dell'informazione*, Codice edizioni, Turín.

Gascón Marcés, A. (2020), Derechos humanos e inteligencia artificial. En Pérez Miras, A., Teruel Lozano, G.M., Raffiotta, E.C., Iadicco, M.P. (dirs.), Romboli, S. (coord.), *Setenta años de Constitución Italiana y cuarenta años de Constitución Española* (pp. 335-350). Madrid: Agencia Estatal Boletín Oficial del Estado.

Grupo de expertos de alto nivel sobre inteligencia artificial (2019), *Directrices éticas para una IA fiable*, Bruselas: Comisión Europea.

Han, B.-C. (2014), *Psychopolitik*, Frankfurt am Main: S. Fischer Verlag GmbH; trad. cast. *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*, Barcelona: Herder.

Han, B.-C. (2020), "La emergencia viral y el mundo de mañana", *El País*, 21 marzo, https://elpais.com/ideas/2020-03-21/la-emergencia-viral-y-el-mundo-de-manana-byung-chul-han-el-filosofo-surcoreano-que-piensa-desde-berlin.html

Harari, Y. N. (2016), *Homo Deus: a Brief History of Tomorrow*, London: Vintage, 2015; trad. cast. *Homo Deus. Breve historia del porvenir*,Barcelona: Debate.

Harari, Y.N. (2020), The World after Coronavirus. *Financial Times*, 20 de marzo de 2020.

Kurzweil, K. (2005), *The Singularity is Near: When Humans Trascend Biology*, Nueva York: Penguin.

Nieva Fenoll, J. (2018), *Inteligencia artificial y proceso judicial*, Madrid: Marcial Pons.

O'Neil, C. (2018), *Armas de destrucción matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia,* Madrid: Capitán Swing; ed. or. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown.

Pietropaoli, S. (2021), "Da cittadino a *user*. Capitalismo, democracia e rivoluzione digitale", *Quaderni del Laboratorio "Hans Kelsen"*, en prensa.

Poundstone, W. (2005), *El dilema del prisionero*, Madrid: Alianza Editorial.

Rodotà, S. (2018), *Vivere la democracia*, Roma-Bari: Laterza.

Romeo, F. (2020), "Giustizia e predittività. Un percorso dal *machine learning* al concetto di diritto". *Rivista di filosofia del diritto*, IX, 1, 107-124.

Sartor, G. (2020), Intelligenza artificiale e diritto. Introduzione. *Rivista di filosofia del diritto, Journal of Legal Philosophy* 1/2020, 65-72.

Simoncini, A., Suweis, S. (2019), Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale. *Rivista di filosofia del diritto*, 1, 87-106.

Solar Cayón, I. (2019), *La inteligencia artificial jurídica. El impacto de la innovación tecnológica en la práctica del derecho y el mercado de servicios jurídicos*, Pamplona: Thomson/Aranzadi.

Tamburrini, G. (2020), *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Roma: Carocci editore.

Žižek, S. (2005), Against Human Rights. *New Left review*, 34, 115-131.

Zuboff, S. (2015), Big Other: Surveillance Capitalism end the Prospects of an Information Civilization. *Journal of Information Technology*, 30, 75-89.

Zuboff, S. (2020a), "Capitalismo de la vigilancia". *Política exterior*, 34, 194, 7-12.

Zuboff, S. (2020b), *La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder*, Barcelona: Paidós; ed. or. *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, Naew York: Public Affairs, 2019.