

Medicina personalizada, algoritmos predictivos y utilización de sistemas de decisión automatizados en asistencia sanitaria. Problemas éticos

Iñigo de Miguel Beriain

Investigador Distinguido
G.I. Cátedra de Derecho y Genoma Humano,
Departamento de Derecho Público,
UPV/EHU
IKERBASQUE Research Professor
inigo.demiguelb@ehu.es

Personalized Medicine, Predictive Algorithms and the Use of Automated Decision Systems in Healthcare. Ethical Issues

ISSN 1989-7022

RESUMEN: Vivimos lo que ya es el nacimiento de una nueva era, marcada por el desarrollo de las nuevas herramientas de Inteligencia Artificial, los denominados sistemas de machine learning, basados en algoritmos capaces de aprender por sí mismos. Esta tecnología es sumamente prometedora, ya que promete mejorar sustancialmente muchos aspectos de nuestra vida, entre ellos, la asistencia sanitaria. El uso de la Inteligencia Artificial en este sector permitirá a buen seguro una mejora de nuestra capacidad de diagnóstico y prognosis. No obstante, también plantea grandes problemas éticos y jurídicos. Entre ellos, la posibilidad o no de autorizar decisiones automatizadas o las necesidades de información a los pacientes; problemas relacionados con el diseño y la validación de estos instrumentos; cuestiones relativas al uso de datos personales para su empleo, o la posibilidad de que acaben generando graves discriminaciones sociales. Este texto expone estos problemas y trata de aportar algunas propuestas de soluciones.

ABSTRACT: We live what is already the birth of a new era, characterized by the development of new Artificial Intelligence tools, the so-called machine learning systems, based on algorithms capable of learning by themselves. This technology is extremely promising, as it offers substantial improvements in many aspects of our lives, including health care. The use of Artificial Intelligence in this sector will surely allow an improvement in our capacity for diagnosis and prognosis. However, it also poses major ethical and legal problems. These include the possibility of authorising automated decisions or the need to provide patients with adequate information; problems relating to the design and validation of these instruments; issues relating to the use of personal data for their use, or the possibility that they may end up generating serious social discrimination. This text sets out these problems and tries to come up with some proposals for solutions.

PALABRAS CLAVE: algoritmos, IA en asistencia sanitaria, toma automatizada de decisiones, protección de datos personales

KEYWORDS: algorithms, AI in health care, automatized decision making, data protection regulation

1. Introducción. La importancia de la Inteligencia Artificial en la medicina del futuro

A día de hoy resulta cada vez más frecuente entender la medicina del futuro como una medicina personalizada,¹ un concepto que no sólo no coincide con el de medicina genómica², sino que va mucho más allá de lo que esta ofrece actualmente. La medicina personalizada podría definirse como “brindar al paciente correcto el tratamiento adecuado en su dosis precisa en el momento oportuno”.³ De este modo, tal vez en breve sea posible alcanzar la que ha sido siempre una aspiración ineludible de la práctica médica, a secas: construir respuestas concretas para problemas particulares.

Durante mucho tiempo, alcanzar este objetivo ha resultado extremadamente complejo. Tanto es así que la historia de la medicina es, en buena medida, un lento progreso hacia un mejor conocimiento de cada patología y cada tratamiento, un progreso que ha conocido diferentes momentos históricos. Cabría, incluso, hablar de una serie de etapas progresivas, en las que hemos incorporado a la medicina el uso de la estadística para dar lugar a la medicina estratificada⁴ o las tecnologías de análisis del ADN que permitieron desarrollar la medicina genómica.

Cada una de estas incorporaciones ha ido permitiendo una mejor comprensión de cada caso y de las expectativas que podrían albergarse en su diagnóstico y tratamiento. La Inteligencia Artificial (IA) aplicada a la práctica clínica promete llevar a la medicina per-



sonalizada un paso más adelante, reduciendo las limitaciones a las que ahora mismos nos enfrentamos. La introducción de la estadística en la medicina, por ejemplo, permitió encontrar patrones generalizables de reacción a los tratamientos, lo que, a su vez, impulsó la diferenciación en la aplicación del recurso sanitario a unos grupos poblacionales u otros. Sin embargo, esta forma de prognosis resulta en ocasiones incluso contraproducente cuando nos hallamos ante un paciente que, por sus circunstancias individuales, no responde de la misma forma que el sujeto medio del grupo poblacional al que se le adscribe.⁵ La medicina genómica, por su parte, encuentra considerables taras en lo que respecta a la capacidad de producir modelos predictivos adecuados. Piénsese que sólo es capaz de encontrar explicaciones absolutamente fehacientes a patologías relacionadas con modificaciones en el genoma y difícilmente a todas ellas, dada la complejidad que algunas presentan.⁶ A todo ello por fin hay que añadir que el coste de esta clase de pruebas puede ser prohibitivo en algunos casos y que no todos los sistemas públicos de salud pueden permitirse costear una política adecuada sobre estas bases (Christensen, Dukhovny, Siebert y Green, 2015).

En estas condiciones, la aplicación de la IA en el diagnóstico y el tratamiento de patologías humanas promete un cambio radical en el escenario de la práctica de la medicina. Es cierto, con todo, que lo que resulta una novedad no es tanto el tipo de herramienta como la forma en que se emplea. Y es que la utilización de algoritmos⁷ es una constante en la práctica de la medicina. No obstante, en el momento actual de la ciencia, el desarrollo de la tecnología informática y la capacidad de procesar cantidades ingentes de datos ha dado lugar a la proliferación del *machine learning*, esto es, de sistemas informáticos automatizados capaces de aprender por sí mismos. Y esto sí que encierra un cambio cualitativo con respecto a todo lo que habíamos visto en el pasado.

Gracias a este nuevo desarrollo tecnológico, proliferan ahora nuevos y asombrosos productos. Así, el sistema de diagnosis pronto se beneficiará de herramientas capaces de apoyar efectivamente la tarea del facultativo, mejorando considerablemente la calidad del resultado final de la atención proporcionada. De sobra es conocido en este campo el IBM Watson, un sistema de inteligencia artificial capaz de interpretar adecuadamente preguntas complejas.⁸ De la asociación de IBM con *Quest Diagnosis*, a su vez, surgió IBM Watson Genomics,⁹ que ha desarrollado IBM Watson Oncology, un sistema de diagnóstico y tratamiento oncológico.¹⁰ En este mismo sentido, cabe recordar que en 2016 se anunció que Google DeepMind colaboraría con el NHS británico para construir un *machine learning system* que -se supone- conseguirá reconocer los síntomas de patologías capaces de amenazar la visión con un simple escaneo digital del ojo humano.¹¹ Otros ejemplos, tal vez menos mediáticos, pero conceptualmente similares a los anteriores, son los modelos que permiten complementar adecuadamente al elemento humano en patologías como los accidentes cardiovasculares (Weng, Reys, Kai, Garibaldi y Qureshi, 2017), lo que incluye, por ejemplo, la interpretación automatizada de los datos de un electrocardiograma (Madani, Arnaout y Mofrad, 2018). También cabe aludir a proyectos que trabajan en el campo de la salud mental, como la predicción del suicidio sobre la base del análisis masivo de datos (Cheng, Li, Kwok, Zhu y Yip, 2017).

Por tanto, es posible señalar que la aplicación de la IA a la medicina está permitiendo crear modelos diagnósticos y predictivos apenas imaginables hace unos años. A nuestro juicio, son los modelos predictivos los que aportan la mayor novedad en el campo de la medicina personalizada, ya permiten una forma de predicción que se puede extender a todo tipo de patolo-

gías que muestren algún patrón de relación determinista (obviamente, las asociaciones caóticas quedarían fuera de sus posibilidades) con todo tipo de datos del usuario/paciente, no sólo los que a primera vista ofrecen relevancia clínica. Tanto es así que el escenario más probable de la medicina del futuro viene construido en torno a un modelo en el que los profesionales de la salud disfrutarán de una información integral proveniente de diferentes fuentes, que se integrarán en modelos automatizados de apoyo a la decisión que les asesorarán adecuadamente sobre el curso más adecuado de acción terapéutica para cada paciente.¹²

Ahora bien, esto no significa que esta nueva forma de medicina se halle exenta de problemas, problemas que tienen que ver con la propia esencia del método que da carta de naturaleza a la creación de algoritmos contruidos a partir del uso masivo de datos. Y es que hay que tener presente que la diferencia fundamental entre el tipo de información que utiliza la medicina tradicional y el uso de herramientas inteligentes de predicción de última generación radica en que, si bien ambas encuentran asociaciones deterministas, la primera es capaz de explicitarlas, mientras que la segunda, generalmente, no lo consigue.¹³ Dicho con otras palabras, la medicina tradicional nos da una explicación inteligible sobre por qué un tratamiento concreto funciona o funcionará en un paciente en concreto. Las herramientas de las que hablaremos ahora, por el contrario, sólo son capaces de predecir, pero no de explicar. Con las únicas excepciones de algunos algoritmos particularmente sofisticados,¹⁴ especialmente complejos y de muy elevado coste (Doshi-Velez, Finale and Kortz, 2017), la mayor parte tipo de construcciones producen únicamente asociaciones, pero no explicaciones razonables sobre los vínculos existentes entre la acción y la reacción. De ahí que se pueda decir que generan correlaciones, no cadenas de causalidad.¹⁵ En consecuencia, estas técnicas despiertan un considerable número de interrogantes completamente novedosos que no encuentran fácil respuesta desde el punto de vista de la ética.

2. IA en medicina: los problemas que advendrán

De cara a entender cuáles pueden ser los problemas derivados de la utilización de la IA en el campo de la medicina, tal vez sea lo más razonable acudir a un caso concreto, que puede servir como herramienta eficaz para entender más fácilmente la índole de las cuestiones sometidas a debate.¹⁶ Éste se hizo público en diciembre de 2017, cuando la revista *Radiology* publicaba un artículo describiendo una investigación capaz de afrontar una de las disfunciones más habituales en la prevención y tratamiento del cáncer de mama (Bahl et al., 2017). Las pruebas de detección de esta patología arrojan a menudo como resultado la detección de masas tumorales. De todas ellas, aproximadamente un 70% se clasifican como claramente benignas, mientras que un 20% se consideran malignas. No obstante, hay un porcentaje de casos, que algunos autores llegan a elevar hasta un 14%, que corresponden a las denominadas *high-risk breast lesions (HRLs)* (Eby et al., 2009). La literatura ha reseñado que un 15% de estos casos acaban convirtiéndose en cánceres (Beegan y Offiah, 2017). Ante esta posibilidad, los facultativos proceden con frecuencia a realizar biopsias e incluso intervenciones quirúrgicas que a menudo acaban resultando completamente innecesarias y que, por tanto, causan al paciente un daño físico y psicológico innecesario.¹⁷ De ahí que cada vez resulten más frecuentes los llamamientos a reducir este exceso de intervencionismo atendiendo a las características de cada paciente (Beegan y Offiah, 2017). El problema, por supuesto, es que esto es muy complejo de realizar en la práctica.

A finales de 2017, un equipo de investigación de MIT compuesto por tres mujeres, Manisha Bahl, directora del *Massachusetts General Hospital Breast Imaging Fellowship Program*, la *MIT Professor* Regina Barzilay, y Constance Lehman, profesora en la *Harvard Medical School* y responsable de la *Breast Imaging Division* en el Departamento de Radiología del *Massachusetts General Hospital* desarrollaron una solución sumamente prometedora a este problema. Esta no es sino un algoritmo inteligente que consigue diferenciar entre unas lesiones y otras a partir del uso de una serie de variables que comprenden datos claramente médicos, como los antecedentes familiares, las biopsias pasadas o los informes de la patología, pero también otros más generales, como la etnia o la edad de la menopausia. La fiabilidad de la herramienta se contrastó gracias a su aplicación sobre 335 de estas lesiones. Las pruebas preliminares realizadas de manera selectiva han mostrado una tasa de aciertos del 97% en el diagnóstico de tumores de mama malignos.¹⁸ Gracias a ello, se consiguió reducir el número de intervenciones innecesarias en, al menos, un 30% con respecto a las aproximaciones anteriores.

Como puede apreciarse, esta nueva herramienta servirá, seguramente, para evitar tratamientos innecesarios y para proporcionar alivio a pacientes que lo necesitan. Sin embargo, sería necio negar que también puede dar lugar a enconadas discusiones acerca de cuestiones que no pueden ser tildadas de poco importantes. Así, por ejemplo, ¿cómo garantizaremos que la herramienta es lo suficientemente fiable? ¿Podremos obligar a un paciente a proporcionar datos particularmente sensibles para decidir sobre su tratamiento? ¿Con qué límites? ¿Podremos utilizar datos de su historia clínica, obtenidos para una finalidad concreta, para decidir sobre una cuestión diferente? ¿Podrá él o ella negarse a que esos datos se utilicen, o puede pedir que se borren sus datos? ¿Qué límites han de conocer las recomendaciones otorgadas por estas herramientas de inteligencia artificial? ¿Pueden constituirse en mecanismos automatizados de decisión? ¿O han de estar siempre supeditados al elemento humano? ¿Cómo resolveremos los casos en los que la recomendación de la IA choque con el juicio del facultativo? A responder a muchas de estas preguntas dedicaremos los próximos apartados de este texto, en los que entraremos, por fin, a debatir las cuestiones éticas que surgen a partir del uso de la inteligencia artificial en el ámbito biomédico.

3. Problemas relacionados con la validación de los mecanismos de IA

La primera cuestión que deberíamos considerar desde un punto de vista ético es la de cómo validar la consistencia científica de estos mecanismos, porque sin ese tipo de pruebas, ¿cómo podemos asegurar la seguridad de un paciente sometido a un tratamiento recomendado por la IA? En el caso de los algoritmos diseñados para favorecer el diagnóstico, contamos con la considerable ventaja de que podemos contrastar sus resultados con las evidencias que pueden presentar los facultativos especializados en la sintomatología presente. Más aún, muchas de las normas con las que ahora contamos ofrecen posibles respuestas a esta cuestión.¹⁹

En el caso de algoritmos enfocados a la prognosis, no obstante, este problema resulta más complejo. Hasta donde llega nuestro conocimiento, los procesos de validación de este tipo de herramientas distan de hallarse adecuadamente regulados, lo que añade un elevado grado de incertidumbre a su empleo. Obviamente, el uso de un algoritmo para adivinar lo que va a suceder después de la aplicación de un tratamiento guarda muy pocas similitudes con la realización de un ensayo o, incluso, un estudio clínico, con lo que difícilmente podremos

aplicarles los mismos modos de validación que utilizamos en el caso de estos últimos. Más aún, un facultativo sólo podrá descartar la validez de una herramienta si es capaz de proporcionar una explicación razonable que aconseje un tratamiento alternativo. Así, por ejemplo, si demuestra que la recomendación de la herramienta podría causar un desenlace fatal, será evidente que el diseño del algoritmo resulta fallido. Pero cuando no es posible efectuar este tipo de contraste de manera eficiente, resulta complicado encontrar alternativas eficientes con las que dar respuesta a este reto.

La forma más razonable de chequeo parece situarse en el ámbito de lo retrospectivo: confrontar sus predicciones con la resolución final de un caso clínico ya acontecido, esto es, con la historia de un paciente ya recuperado o fallecido. No obstante, incluso este sistema puede tener sus limitaciones, por varios motivos. En primer lugar, porque las predicciones suelen ser individuales, con lo que lo que ha funcionado bien para unos pacientes no necesariamente lo hará en el caso de otros. Más aún, es posible que en el momento actual haya datos disponibles que no se hubieran recogido en historias antiguas. Por fin, hay que tener presente que los algoritmos, si de verdad son inteligentes, cambian, no son vehículos estáticos. Con lo que lo que resulta cierto a día de hoy puede no serlo ya en el futuro (Senger y O'Leary, 2018). De ahí que tenga mucho sentido pensar en una validación dinámica, que tenga presentes los datos con los que contamos antes de permitir la comercialización de una herramienta, pero que realice un seguimiento atento de la misma después de su introducción en el mercado, tal y como ha propuesto, por ejemplo, la Federal Drug Administration de los EE. UU.²⁰ o el Parlamento Europeo, que ha dictaminado que de cara a asegurar la fiabilidad de estas herramientas serán necesarias "evaluaciones periódicas sobre la representatividad de los conjuntos de datos [así como] examinar la exactitud e importancia de las predicciones".²¹

Teniendo todo esto presente, parece claro que la forma en que validaremos estos instrumentos constituye un problema ético relevante. De ahí, en suma, que sea necesario plantearse la necesidad de regular los mecanismos que garantizarán una adecuada auditoría dinámica de los sistemas de inteligencia artificial de los que ahora nos ocupamos,²² una necesidad que no es fácil de cumplimentar, por cuanto muchos de ellos son inherentemente opacos.²³ Esta situación parece particularmente preocupante si tenemos en cuenta que, en realidad, en el caso de las herramientas de las que nos estamos ocupando, cabe pensar en la necesidad de dos formas de validación diferentes, ambas muy necesarias. La primera es la que se refiere a la dinámica interna del instrumento: es necesario certificar que en su confección no se han introducido voluntaria o involuntariamente sesgos que alteren su funcionamiento²⁴ de acuerdo con los intereses o las inclinaciones de los programadores.²⁵ La segunda, en cambio, se refiere a la precisión de su análisis, a la exactitud de sus predicciones, lo que muchas veces dependerá de la propia calidad de los datos utilizados en su confección.²⁶ Nos hallamos, en suma, ante un escenario complejo, con múltiples aristas, que necesita de soluciones urgentes, pero que no son fácilmente diseñables. La situación, además, se agrava si tenemos en cuenta lo que indicaremos en el próximo apartado.

4. Problemas en la creación de algoritmos: diseño de los algoritmos inteligentes

Un problema de excepcional importancia a añadir al que acabamos de mostrar es el que se deriva de un hecho difícilmente alterable: que la utilidad de los algoritmos depende de la capacidad

de predicción que les proporciona el uso de un número elevado de variables con incidencia en el resultado que se pretende pronosticar. Un algoritmo utilizará todo tipo de variables que parezcan ofrecer alguna correlación determinista con respecto al acontecimiento sobre el que se pretende ofrecer un pronóstico, con independencia de cuáles sean esas variables o el tipo de dato del que se parta cuando se empleen. Y esto en ocasiones incluirá datos como el género, etnia, profesión, ideología, tendencia sexual u otros de índole similar, que gozan de una protección singular, de acuerdo con lo dispuesto en el artículo 16.2 de la Constitución Española o con el artículo 9.2.a) del Reglamento (UE) 2016/679 o el artículo 9.1 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

En tales circunstancias, puede producirse una tensión inevitable entre la correlación estadística posible entre variables de este tipo y el suceso a determinar y el derecho recogido en la mayor parte de los ordenamientos constitucionales de no discriminar a nadie en función de determinantes como los descritos. Pensemos en un caso en el que un algoritmo excluye el acceso a un tratamiento a una persona, y un análisis muestra que la etnia de pertenencia juega un papel fundamental en este sentido. Es más que posible que pensemos que esto sería una discriminación injusta, aunque también es cierto que si, efectivamente, la etnia es un factor que explica la futilidad de un tratamiento no hablaríamos ya de discriminación (por cuanto no habría injusticia). Pero, ¿y si no tenemos constancia de que, efectivamente, haya un motivo razonable por el que ligar esa variable a la recomendación final? Más aún, ¿sobre qué base legal le habremos solicitado al paciente un dato de este tipo si no tenemos constancia de su relevancia médica? El escenario resulta, desde luego, complejo. En estas circunstancias, nos tememos, la posibilidad de que esa discriminación se produzca finalmente o no dependerá de cuál sea el efecto de la recomendación, el ámbito de acción que se confiera a un algoritmo, el derecho del que disponga el usuario a la hora de denegar el acceso a la información de este tipo y otros muchos factores que habría que ponderar.

Por tanto, la conclusión a la que podríamos llegar es que no resulta de por sí censurable que un algoritmo utilice variables sensibles como la religión, ideología, género, etnia, tendencia sexual, etc. Sin embargo, el sistema de salud debería posibilitar al usuario negarse a proporcionar esos datos y evitar el uso de los que ya constan a disposición del responsable del tratamiento, pero para cuyo uso específico para la finalidad que nos ocupa no ha sido proporcionado consentimiento explícito. Más aún, el responsable debería informar siempre al paciente/usuario de su derecho a negarse a que se empleen ese tipo de datos, si no hay un motivo médico solvente por el que solicitarlos, sobre la base de los derechos que le proporciona la ley. Ahora bien, también tendría, nos tememos, que informarle de las consecuencias que ello podría acarrear de cara al grado de precisión a alcanzar mediante el uso del dispositivo de inteligencia artificial o, incluso, la imposibilidad de emplear este mecanismo. Todo ello encaminado, de un lado, a evitar que el uso de estos datos derive en una discriminación injusta entre las personas sobre la base de su etnia, género, posición social, etc., pero también a que un paciente no pueda beneficiarse de una predicción más atinada por un exceso de protección por parte del personal sanitario.²⁷

La cuestión, no obstante, se complica aún más si tenemos presente que algunos algoritmos pueden utilizar estos datos implícitamente, esto es, inferir algunos de ellos de otros con los que ya se cuenta, e introducirlos como variables en su mecanismo de procesamiento.²⁸ De este modo, el sujeto podría ver violentada su negativa a proporcionar permiso para recabar estos da-

tos sin que existiera una evidencia palpable de la transgresión. Y, lo que es peor, pudiera ser que esto sucediese sin que ni siquiera el desarrollador del algoritmo fuera plenamente consciente de todo ello.²⁹ La cuestión, por supuesto, radica en cómo conseguir adivinar si, efectivamente, un algoritmo utiliza o no esta clase de datos. El recurso a predicciones contra fácticas, esto es, la alteración de algunos de los datos sensibles para contrastar la solución que el programa ofrece ante estas alternativas puede ser un medio eficiente para ello. Habrá que ver cuáles pueden ser sus limitaciones en la práctica (Wachter, Mittelstadt y Russell, 2018).

5. Problemas relacionados con la toma automatizada de decisiones

Otra cuestión esencial que suscita la AI aplicada a la práctica clínica es bajo qué condiciones o, mejor aún, con qué límites podemos amparar su utilización a la hora de efectuar un diagnóstico o determinar un tratamiento. Se trataría en este caso de delimitar con precisión el poder de decisión que podemos llegar a conceder a la herramienta. Imaginemos, por ejemplo, que en el caso del algoritmo descrito en apartados anteriores, el sistema recomienda no intervenir en una paciente que muestra una lesión del tipo de las que un profesional sanitario habitualmente optaría por reseca. ¿Qué validez tendría la recomendación del sistema de IA? ¿Debería imponerse su criterio al de un facultativo? ¿Debería ser el paciente quien optara sobre la base de una información precisa?

Esta es, sin duda, la cuestión crucial. Hay que tener presente a este respecto que el abanico de opciones es amplio, por cuanto podemos dotar a estos dispositivos de un papel de mero apoyo a las decisiones facultativas a otro en el que el personal sanitario tenga que justificar la adopción de una decisión divergente con respecto a lo que sugiere la máquina. Más aún, cabría incluso dotar al sistema informático de un papel protagonista, en el que sería el propio algoritmo el que decidiera acerca del tratamiento o las medidas de prevención o cuidado a proveer al usuario/paciente, con total independencia del elemento humano.³⁰

Esta última opción resulta hoy descartable por motivos técnicos, éticos, organizacionales y, por encima de todo, jurídicos, al menos en el marco de la UE. Con respecto a lo primero, hay que recordar que los mecanismos con los que ahora mismo contamos son todavía un tanto rudimentarios y su fiabilidad insuficiente, como han demostrado algunas experiencias con el sistema Watson de IBM.³¹ A nuestro juicio, su progresiva implementación y la validación que irá otorgándoles su uso contribuirán a paliar este problema, pero en estos momentos sería prematuro concederles un poder de esta índole. A ello hay que añadir, desde una esfera organizacional, la dificultad que entrañaría vaciar de funciones decisorias en torno a las medidas a adoptar al personal sanitario, por no mencionar la cuestión de quién asumiría la responsabilidad jurídica por un fallo en un sistema de toma de decisiones si estas dependen exclusivamente de un procedimiento de IA. De ahí que parezca perfectamente razonable la tendencia jurídica actual, que apunta a la prohibición de las decisiones automatizadas, reflejada, por ejemplo, en el Reglamento Europeo de Protección de Datos.³²

Ahora bien, dicho esto hay que referir que es muy complejo especificar cuál ha de ser la magnitud de la intervención del elemento humano mínima y necesaria para considerar que la decisión no se basa únicamente en el tratamiento automatizado. Y esto es una inconcreción sumamente disfuncional que introduce una amplia esfera de incertidumbre en nuestra regulación jurídica,

que no se ha producido casualmente. En principio, todo paciente o usuario tiene derecho un derecho a obtener una información precisa acerca de cómo se ha adoptado una decisión relativa a su salud.³³ La cuestión fundamental radica en saber en qué consiste ese derecho a la información, esto es, qué requisitos han de cumplirse para determinar si un paciente ha sido o no adecuadamente informado no ya del uso de una herramienta de inteligencia artificial en el proceso clínico, sino de cómo funciona. Y en este contexto, no está claro si la idea de conocer se circunscribe únicamente a saber si y para qué se utiliza un mecanismo de inteligencia artificial, o cuáles serán las posibles consecuencias de su uso, sino si también habría que incluir que la necesidad de dotar al paciente de una información precisa sobre la lógica del mecanismo.

A este respecto, se han proporcionado diversas propuestas. Así, por ejemplo, Wachter, Mittlstadt y Floridi (2017), consideran que una explicación es una descripción que permite explicar de forma inteligible para un ser humano cómo funciona un proceso por el que un encargado de toma de decisiones consigue alcanzar una determinada conclusión gracias al empleo de una serie de datos. Doshi-Velez et al. (2017) han especificado que esto implica que la explicación es capaz de dotar de adecuada respuesta a cuatro preguntas fundamentales: ¿Cuáles fueron los principales factores que incidieron en la decisión? ¿Habría cambiado la decisión si hubiésemos alterado alguno de ellos? ¿Por qué dos casos que parecían muy similares obtuvieron diferente consideración o viceversa? La Guía de Buenas Prácticas elaborada por el Grupo del Artículo 29, por fin, dice textualmente que *“en lugar de proveer una explicación matemática compleja sobre cómo funcionan los algoritmos o el machine-learning, el responsable del tratamiento de los datos debería considerar la posibilidad de utilizar formas claras y completas de proporcionar información al sujeto, por ejemplo: las categorías de datos que se han o van a ser utilizados en la elaboración de perfiles o los procesos de toma de decisiones; por qué estas categorías se consideran pertinentes; cómo se construye un perfil utilizado en el proceso de decisión automatizado; y cómo se utiliza para tomar una decisión referente al sujeto”*.

El mayor problema que habremos de afrontar, sea cual sea el sentido del que dotemos a la idea de información, es que en ocasiones es extremadamente complejo, incluso para el desarrollador de un algoritmo, proporcionar una respuesta clara a todos los interrogantes que un paciente o un facultativo puedan plantear, debido a la propia complejidad de estas herramientas, como ya hemos señalado anteriormente. Este obstáculo podría solventarse utilizando sofisticadas herramientas informáticas, sistemas de inteligencia artificial capaces de proporcionar explicaciones sobre el funcionamiento de los algoritmos, de los que ya hemos hablado en apartados anteriores. La dificultad que esta opción encierra, no obstante, es que ello encarecería considerablemente los costes, lo que acabaría expulsando de este sector a las PYMES y abriendo la puerta a comportamientos de tipo monopolístico, que no parecen lo más convenientes (Doshi-Velez et al., 2017).

El recurso a la intervención de un recurso externo y neutral parece, de nuevo, el más eficiente de cara a garantizar al paciente, cuando menos, la capacidad de obtener las respuestas a las preguntas que pueda plantearle el uso de la IA.³⁴ Podría, por supuesto, pensarse en los comités de ética asistencial como medios con los que satisfacer esta necesidad, pero tal vez esto sería demasiado aventurado. No debemos olvidar que, a fin de cuentas, muchos de sus miembros no poseen cualificación precisa en IA, ni disponibilidad de tiempo y recursos para ofrecer un asesoramiento del tipo que esta forma de concebir la asistencia traería consigo. Por eso mismo, será probablemente necesario recurrir a otros mecanismos, que habrá que diseñar

en breve. En este sentido, nos resulta particularmente interesante la propuesta realizada por Fiske et al. (2018) de institucionalizar una nueva figura, lo que denominan *Health Information Counselors*, que podrían caracterizarse como profesionales particularmente cualificados en el funcionamiento de la IA y capaces de ofrecer respuestas razonables sobre la tecnología a aplicar tanto a pacientes como a profesionales sanitarios.

6. Problemas relacionados con el uso de datos

La utilización de modelos diagnósticos o terapéuticos contruidos sobre la base del uso de datos personales del paciente o el usuario de la sanidad abre, de otro lado, un nuevo desafío a la construcción jurídica de la defensa de su intimidad. En términos generales, la mayor parte de los ordenamientos jurídicos que se han ocupado de esta cuestión dividen el tratamiento de datos de salud en dos grandes momentos, el de su obtención y utilización para fines diagnósticos y terapéuticos y el de su uso para fines de investigación. De ahí que frecuentemente para el empleo de datos en investigación sea necesario el consentimiento explícito del paciente, un consentimiento diferenciado del que se otorga para su uso clínico.

En principio, sin embargo, se considera que no tiene en cambio sentido pensar en un consentimiento diferenciado para el uso de datos para el diagnóstico y el tratamiento. Esta creencia generalizada se ha construido sobre la base de que los datos que permiten diagnosticar también avalan la decisión clínica de adoptar un tratamiento u otro. Sin embargo, este panorama puede cambiar radicalmente en el momento en que los algoritmos empiecen a influir en la toma de decisiones sobre el diagnóstico. En algunos casos, no será así, por supuesto, ya que las herramientas no emplearán otros datos que los que figuran en las historias de los pacientes/usuarios o los que se han obtenido para diagnosticar sus patologías.³⁵

En otras ocasiones, sin embargo, la utilización de los algoritmos resulta mucho más problemática, por cuanto éstos utilizan datos que van mucho más allá de los que se requieren para el diagnóstico. Pensemos, sin ir más lejos en el caso del algoritmo del MIT expuesto. Un diagnóstico de cáncer de mama requiere un número de pruebas relativamente escasa. Es sólo cuando se sospecha de la existencia del tumor cuando se empiezan a recabar más, y esa recolección puede ser exhaustiva si hacemos intervenir a un algoritmo. Piénsese que el que citamos utiliza, además, de la información que habitualmente se extrae de una mamografía, datos como la edad del primer embarazo, la de la menopausia, la de la primera menstruación, la etnia, etc. Y todo ello para realizar una recomendación de tratamiento que en la mayor parte de los casos implicará un gran beneficio para el paciente, pero que en algunos casos puede suponerle un riesgo vital que las opciones actuales eliminan.³⁶

De ahí que pueda suceder que un paciente que conozca -y la normativa jurídica en general obliga a informarle- de la finalidad de la recogida de datos, de la intervención de la herramienta y de los riesgos asociados pueda negarse a proporcionarlos para obligar al sistema a proporcionarle una opción terapéutica que es más cara y mucho menos recomendable para la mayoría de los pacientes.³⁷ Obviamente, esta voluntad de impedir la utilización de un mecanismo como el que describimos, incluso como mero apoyo a la decisión facultativa, puede ser mucho mayor en casos como el que hemos descrito, por cuanto muchos pacientes temerán que pueda utilizarse para negarles un tratamiento, en lugar de interpretar que lo que la

herramienta pretende es evitar intervenciones innecesarias y favorecer la implementación de cuidados paliativos.

La cuestión, por tanto, consiste en si un paciente podría, caso de requerírsele que proporcionase una batería de datos personales adicionales en un momento posterior al diagnóstico, negarse a colaborar con el sistema. Y, en principio, la solución a esta cuestión es compleja, por cuanto provoca una colisión entre diferentes tipos de deberes. De un lado, asumimos habitualmente que un paciente tiene la obligación de proporcionar al sistema de salud información fiable sobre su salud.³⁸ De otro, no obstante, tendemos a garantizar a cualquier persona que nadie podrá emplear sus datos para otros fines que para aquellos para los que preste su consentimiento, cierto que con algunas excepciones, como las referentes al cuidado de la salud pública, por ejemplo. Cómo resolvamos en el futuro la fricción entre ambos tipos de deberes definirá en buena medida cómo serán las sociedades en las que la IA se implemente en el campo de la medicina.

7. Problemas relacionados con la discriminación de colectivos específicos

Debemos, por fin, tener presente que la introducción de sistemas de IA en la asistencia sanitaria implica también problemas de justicia social, relacionados con la posible discriminación de colectivos enteros. Y es que, como señala Elena Gil González (2016, 124) *“La categorización puede ser una herramienta muy útil, pero también entraña riesgos de cometer errores al conectar con una persona determinadas características con ciertos comportamientos. Por su parte, también hay riesgo de que los perfiles basados en características como raza, etnia o religión creen estereotipos poco precisos que causen discriminación”*. La raíz del problema tiene que ver, simplemente, con el tipo de datos sobre los que se construyen ahora mismo la mayor parte de los algoritmos que se están utilizando en asistencia sanitaria: datos procedentes de historias clínicas recabadas en un contexto en que la población era mayoritariamente de una etnia homogénea. Esto es un problema que ya se ha detectado en la aplicación de la medicina genómica. Cuando intentamos aplicar las herramientas disponibles a poblaciones minoritarias se hace patente que estas resultan mucho menos eficaces, ya que los estudios genéticos están sesgados de modo cuasi-irreparable por el factor étnico (Popejoy y Fullerton, 2016). Y, además, sin que haya mucho que podamos hacer al respecto, salvo que intentemos compensar esta deficiencia de nuestra investigación en genética realizando estudios más amplios que integren todas las etnias de una manera mucho más proporcionada (Martin et al., 2017). Es más que probable que, al menos en un primer momento, tengamos que asumir que las herramientas de IA reproducirán este problema. Pero, de ser así, será muy complejo garantizar una igualdad real en el acceso a los recursos reales del sistema sanitario.

Bibliografía

Agencia Española de Protección de Datos (AEPD) y a la Asociación Española para el Fomento de la Seguridad de la Información, ISMS Forum Spain. *Código De Buenas Prácticas En Protección De Datos Para Proyectos Big Data*, en: http://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/2017/Guia_Big_Data_AEPD-ISMS_Forum.pdf. Última consulta: 18 de febrero de 2019.

- Article 29 Data Protection Working Party. "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679", 17/EN WP251rev.01. Adopted on 3 October 2017 As last Revised and Adopted on 6 February 2018. En: https://iapp.org/media/pdf/resource_center/W29-auto-decision_profiling_02-2018.pdf. Última consulta: 18 de febrero de 2019.
- Beegan, A.A., & Offiah, G. (2017). "High-Risk Breast Lesions". In *Breast Imaging*. <https://doi.org/10.5772/intechopen.70616>
- Bahl, M., Gaffney, S., McCarthy, A. M., Lowry, K. P., Dang, P. A., & Lehman, C. D. (2017). "Breast Cancer Characteristics Associated with 2D Digital Mammography versus Digital Breast Tomosynthesis for Screening-detected and Interval Cancers". *Radiology*, 287(1), 49-57. <https://doi.org/10.1148/radiol.2017171148>
- Berberich, Amanda J., Ho, R. & Hegele, R. A. (2018). "Whole genome sequencing in the clinic: empowerment or too much information?". *CMAJ*, February 05, 2018, 190(5), E124-E125.
- Boyd, Danah and Crawford, K. (2011). "Six Provocations for Big Data". *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. En: <https://ssrn.com/abstract=1926431> o <http://dx.doi.org/10.2139/ssrn.1926431>. Última consulta: 18 de febrero de 2019.
- Cesuroglu, T., Van Ommen, B., Malats, N., Sudbrak, R., Lehrach, H., & Brand, A. (2012). "Public health perspective: from personalized medicine to personal health". *Personalized Medicine*, 9(2), 115-119. <https://doi.org/10.2217/pme.12.16>
- Cheng Q., Li T.M., Kwok C.L., Zhu T., Yip P.S. (2017). "Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study". *Journal of Medical Internet Research*. 19(7): e243
- Christensen K.D., Dukhovny D., Siebert U., Green R.C. (2015). "Assessing the Costs and Cost-Effectiveness of Genomic Sequencing". *Journal of Personalized Medicine*. 5(4): 470-486.
- Cotino, L. (2007). "Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales". *Dilemata*, nº 24, 2017, 131-150.
- Doshi-Velez, Finale, and Mason Kortz. (2017). "Accountability of AI Under the Law: The Role of Explanation". *Berkman Klein Center Working Group on Explanation and the Law*, Berkman Klein Center for Internet & Society working paper. Available at SSRN: <https://ssrn.com/abstract=3064761> or <http://dx.doi.org/10.2139/ssrn.3064761>. Última consulta: 18 de febrero de 2019
- Eby, P. R., Ochsner, J. E., DeMartini, W. B., Allison, K. H., Peacock, S., & Lehman, C. D. (2009). "Frequency and upgrade rates of atypical ductal hyperplasia diagnosed at stereotactic vacuum-assisted breast biopsy: 9-versus 11-gauge". *AJR. American Journal of Roentgenology*, 192(1), 229-234. <https://doi.org/10.2214/AJR.08.1342>
- European Science Foundation, Personalized Medicine for the European Citizen. Towards more precise medicine for the diagnosis, treatment and prevention of disease (iPM), p. 13. En: <http://archives.esf.org/coordinating-research/forward-looks/biomedical-sciences-med/current-forward-looks-in-biomedical-sciences/personalised-medicine-for-the-european-citizen.html>. Última consulta: 12 de febrero de 2019
- Fiske, Amelia, Buyx, Alena, Prainsack, Barbara, Health Information Counselors. (2018). "A New Profession for the Age of Big Data". *Academic Medicine*. Volume Publish Ahead of Print - Issue doi: 10.1097/ACM.0000000000002395
- Genevès, P., Calmant, T., Layaïda, N., Lepelley, M., Artemova, S., & Bosson, J.L. (2017). "Predicting At-Risk Patient Profiles from Big Prescription Data". Recuperado de <https://hal.inria.fr/hal-01517087>.
- Gil González, E. (2016). *Big data, privacidad y protección de datos*. Agencia Española de Protección de Datos y Agencia Estatal Boletín Oficial del Estado, Madrid
- International Consortium for Personalised Medicine, Action Plan, 2017, accessible en: https://www.icpermed.eu/media/content/ICPerMed_Actionplan_2017_web.pdf. Última consulta: 18 de febrero de 2019.
- Madani, A., Arnaout, R. and Mofrad, M. (2018). "Fast and accurate view classification of echocardiograms using deep learning". *Npj Digital Medicine*, 1(1), 6. <https://doi.org/10.1038/s41746-017-0013-1>.

- Martin A.R., Gignoux C.R., Walters R.K., Wojcik G.L., Neale B.M., Gravel S., Daly M.J., Bustamante C.D, Kenny E.E. (2017). "Human demographic history impact genetic risk prediction across diverse populations". *Am. J. Hum. Genet.* 100, 635–649.
- Mirnezami, R., Nicholson, J., & Darzi, A. (2012). "Preparing for precision medicine". *The New England Journal of Medicine*, 366(6), 489-491. <https://doi.org/10.1056/NEJMp1114866>
- McGinnis, J. M., Williams-Russo, P. and Knickman, J.R. (2002). "The case for more active policy attention to health promotion". *Health Affairs*, 21(2): 78-93.
- Popejoy, A.B. & Fullerton, S.M. (2016). "Genomics is failing on diversity". *Nature*, 538, 161-164.
- Redekop W.K. & Mladi D. (2013). "The faces of personalized medicine: a framework for understanding its meaning and scope". *Value Health*, 16(6). S4-S9.
- Reuter, M.S. et al. (2018). "The Personal Genome Project Canada: findings from whole genome sequences of the inaugural 56 participants". *CMAJ*, 190 (5). Disponible en: <http://www.cmaj.ca/content/190/5/E126>. Última consulta: 12 de febrero de 2019.
- Senger, J. M., & O'Leary, P. (2018). "Big data and human medical judgment regulating next-generation clinical decision support". In I. G. Cohen, H. Fernandez Lynch, E. Vayena, & U. Gasser (Eds.), *Big data, health law, and bioethics*. Cambridge, England: Cambridge University Press, 283-294.
- Surden, H. (2017). "Values Embedded in Legal Artificial Intelligence". *U of Colorado Law Legal Studies*. Research Paper No. 17-17. En: <https://ssrn.com/abstract=2932333> or <http://dx.doi.org/10.2139/ssrn.2932333>. Última consulta: 18 de febrero de 2019.
- Trusheim M.R., Berndt E.R. & Douglas F.L. (2007). "Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers". *Nat Rev Drug Discov.* 6(4), pp. 287-93
- US Food and Drug Administration (2015). "Balancing Premarket and Postmarket Data Collection for Devices Subject to Premarket. Approval Guidance for Industry and Food and Drug Administration Staff". Disponible en: <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm393994.pdf>. Última consulta: 21 febrero de 2019.
- UK Information Commissioner's Office (ICO) (2017). *Big data, artificial intelligence, machine learning and data protection* (p. 86). Disponible en: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>. Última consulta: 18 de febrero de 2019
- Wachter, S., Mittelstadt, B. & Floridi, L. (2017). "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation". *International Data Privacy Law*, pp. 76-99.
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR". Disponible en arXiv:1711.00399 [cs].
- Weng S.F., Reps J., Kai J., Garibaldi J.M. & Qureshi N. (2017), "Can machine-learning improve cardiovascular risk prediction using routine clinical data?". *PLOS ONE*, 12(4).

Notas

1. A veces, no obstante, se habla de medicina de medicina de precisión en lugar de medicina personalizada, entendiéndose como tal un sistema basado en múltiples fuentes de información que sirven adecuadamente a las necesidades de la clínica y la investigación (Mirnezami, Nicholson y Darzi, 2012). A veces resulta, en realidad, muy complejo trazar diferencias entre ambos conceptos y son muchos los documentos que los consideran prácticamente sinónimos. Resulta, no obstante, innegable que la idea de medicina de precisión tiene más raigambre en Estados Unidos, mientras que en la UE se prefiere el concepto de medicina personalizada (Cesuroglu et al., 2012). Quien quiera entender por qué en EE. UU. se produjo el cambio terminológico puede consultar una interesantísima información proporcionada por el NIH en la siguiente página web: <https://www.google.pl/search?q=what+is+the+difference+between+precision+medicine+and+personalized+me>

dicine&oq=what+is+the+difference+between+precision+medicine&aqs=chrome.1.69i57j0.8814j1j4&sourceid=chrome&ie=UTF-8. Última consulta: 12 de febrero de 2019.

2. Como apunta el Informe de la European Science Foundation, "Así, muchos consideran que la medicina personalizada es sinónimo de medicina genómica. Desafortunadamente, esta perspectiva crea una serie de problemas. Primero, es demasiado estrecha en su visión. El rango de información que posiblemente se pueda usar para personalizar la asistencia sanitaria es mucho mayor que los datos de la secuencia del genoma. En consecuencia, una visión a largo plazo de un verdadero enfoque basado en sistemas para la salud se vería obstaculizado por la percepción de que la medicina personalizada se basa simplemente sobre la predicción basada en la secuenciación del genoma individual. Además, la receptividad de los interesados en el desarrollo y la implementación de la medicina personalizada podría resultar influenciado negativamente por la creciente percepción de que la genómica no está a la altura de las expectativas y aún no ha cumplido la promesa de explicar la susceptibilidad a toda una serie de enfermedades. De este modo es como lograremos estar en condiciones de responder a las críticas que se hacen a la medicina personalizada sobre la base de la idea errónea de que ésta se refiere únicamente a la predicción genómica" (p. 13).
3. Obviamente, hay múltiples definiciones de medicina personalizada, lo que a veces crea cierta confusión terminológica (Redekop y Mladi, 2013).
4. Como tal puede entenderse la identificación de subgrupos de pacientes con una enfermedad en particular que responden de la misma forma a una medicina en particular o sufren los mismos riesgos de padecer un efecto secundario al afrontar un determinado tratamiento (Trusheim, Berndt y Douglas, 2007).
5. Así, por ejemplo, puede resultar que no proporcionemos un tratamiento concreto a un paciente por pertenecer al grupo de los trasplantados de riñón cuando, en realidad, en su caso concreto resultaría perfectamente pertinente por sus características particulares. Lo mismo podría ocurrir, por exponer otro caso, en menores que responden a una medicina de manera diferente a lo que es habitual en su edad, o de sujetos que, aun perteneciendo a una etnia particularmente poco sensible a los efectos secundarios de un medicamento, los sufren particularmente por sus circunstancias personales.
6. Prueba palpable de estas disfunciones es que en el reciente estudio publicado en relación con el *Personal Genome Project*, que ha revelado que las estimaciones realizadas a partir del análisis del ADN de los participantes a menudo diferían de lo que mostraba su historial clínico. Como han escrito Amanda J. Berberich, Rosettia Ho y Robert A. Hegele (2018), "Reuter y sus colegas descubrieron que ninguna de las mutaciones "probablemente patógenas" identificadas entre los participantes se asoció con una enfermedad clínicamente manifiesta en el momento del diagnóstico genético. En ausencia de manifestaciones clínicas, y sin directrices o instrucciones sobre cómo tratar clínicamente este tipo de información, la divulgación de dicha información genética a los participantes o al personal sanitario que los atiende podrían potencialmente tener tanto aspectos positivos como consecuencias negativas. Subestimar el riesgo podría provocar una falsa sensación de seguridad, pero sobreestimarla podría crear una preocupación innecesaria" (p.124-125). Tal vez todo ello tenga que ver con las estimaciones realizadas por McGinnis Williams-Russo y Knickman (2002) en un estudio realizado hace unos años, de acuerdo con el cual las circunstancias sociales y el comportamiento de las personas influían más en su salud que su carga genética.
7. La definición básica de un algoritmo es simple: un conjunto de instrucciones paso a paso para resolver un problema o realizar una tarea. Una receta para galletas con trocitos de chocolate es un algoritmo. Obviamente, el tipo de algoritmos de los que hablaremos aquí es un poco más complejo. Acerca de qué algoritmos se están utilizando en el campo de la salud, véase: 10 Algorithms That Are Changing Health Care *Algorithms for Innovation*. En: <http://uofuhealth.utah.edu/innovation/blog/2015/10/10AlgorithmsChangingHealthCare.php>. Última consulta: 12 de febrero de 2019.
8. Véase al respecto: <http://www-03.ibm.com/marketing/mx/watson/what-is-watson/>. Última consulta: 18 de febrero de 2019.
9. Véase: <https://www.ibm.com/watson/health/oncology-and-genomics/genomics/>. Última consulta: 18 de febrero de 2019.

10. El Libro Blanco de WFG está disponible en: <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=HLW03064USE&>. Última consulta: 18 de febrero de 2019.
11. Google DeepMind pairs with NHS to use machine learning to fight blindness, The Guardian, 5 julio 2016, accessible en: <https://www.theguardian.com/technology/2016/jul/05/google-deepmind-nhs-machine-learning-blindness>. Última consulta: 18 de febrero de 2019.
12. Véase: International Consortium for Personalised Medicine, Action Plan, 2017, accessible en: https://www.icpermed.eu/media/content/ICPerMed_Actionplan_2017_web.pdf. Última consulta: 18 de febrero de 2019. En el mismo sentido: ESF, Personalized Medicine for the European Citizen. Towards more precise medicine for the diagnosis, treatment and prevention of disease (iPM), cit., p. 21.
13. Como escribe Cotino (2017, 131-150), “Frente a la contrastación de una hipótesis a partir de los datos, se descubren correlaciones sin conocer previamente la causa. Así sucede al probar casi aleatoriamente la posible correlación entre datos en principio totalmente distantes (¿Compra distintos alimentos la gente en función del estado del clima? ¿Cómo influye el embarazo en las decisiones de consumo? ¿Sería posible ofrecer seguros de salud en función de las búsquedas en Google o del análisis de las preferencias alimentarias manifestadas por los usuarios y por sus redes de amigos en espacios sociales de internet?)”.
14. Así, por ejemplo, “el programa Watson Paths, concretamente, permite que los médicos vean las pruebas y el sendero de deducciones que ha seguido el ordenador para lanzar sus conclusiones (por ejemplo, cómo ha concluido el diagnóstico médico que realiza). Este experimento pionero intenta dar una solución a esta falta de supervisión humana de los resultados de los algoritmos, a la falta de una traducción máquina-humano, que sin duda seguirá avanzando a medida que la ciencia del big data progrese. El ordenador Watson no es la única iniciativa en este sentido. Algunas empresas aseguran que sus empleados revisan las recomendaciones que hacen sus ordenadores, «aunque es raro que rechacen lo que dictan los algoritmos”. Y es que, como señala Elena Gil González (2016, 43), “La categorización puede ser una herramienta muy útil, pero también entraña riesgos de cometer errores al conectar con una persona determinadas características con ciertos comportamientos. Por su parte, también hay riesgo de que los perfiles basados en características como raza, etnia o religión creen estereotipos poco precisos que causen discriminación”.
15. Como indica Elena Gil González (2016, 28) “se está produciendo un cambio de la «causalidad» a la «correlación». Ya no importa tanto descubrir la causalidad entre dos hechos, sino su correlación. Así, en lugar de tratar de comprender exactamente por qué una máquina se estropea o por qué los efectos secundarios de un medicamento desaparecen, el big data permite a los investigadores recoger y analizar cantidades masivas de datos sobre estos eventos y todo lo asociado a éstos, para encontrar relaciones entre variables que permitan descubrir patrones ocultos y predecir cuándo estos eventos pueden volver a ocurrir”.
16. Nótese que hemos traído estos ejemplos a colación exclusivamente por su capacidad para ayudarnos a exponer la problemática a la que ahora nos enfrentamos. No es nuestra misión certificar la utilidad real de estas herramientas. Incluso aunque habláramos de casos ficticios, seguirían mostrando la misma utilidad para el análisis teórico. Si hemos optado por casos extraídos de la realidad es sólo para mostrar que no estamos hablando de una hipótesis muy lejana en el tiempo o muy poco factible en la práctica clínica.
17. De acuerdo con la información disponible en Salud Digital, “los médicos manejan la patología de diferentes maneras. Algunos proceden a la intervención en todos los casos, mientras que otros la reservan solo para las lesiones que tienen índices de cáncer más altos, como “hiperplasia ductal atípica” (ADH) o “carcinoma lobular in situ” (LCIS). El primer enfoque requiere que el paciente se someta a una cirugía dolorosa, lenta y costosa que generalmente es innecesaria; el segundo enfoque es impreciso y podría dar lugar a tumores faltantes en lesiones de alto riesgo distintas de ADH y LCIS” (información disponible en su página web, https://www.consalud.es/saludigital/86/el-mit-desarrolla-una-inteligencia-artificial-contra-el-cancer-de-mama_43572_102.html. Última consulta: 18 de febrero de 2019).
18. Como señala Salud Digital, “probado en 335 casos de alto riesgo, el modelo detectó correctamente el 97% de los cánceres de mama como malignos y redujo el número de intervenciones en pacientes con tumores benignos en más del 30% en comparación con los enfoques existentes” (información disponible en su página web, https://www.consalud.es/saludigital/86/el-mit-desarrolla-una-inteligencia-artificial-contra-el-cancer-de-mama_43572_102.html. Última consulta: 12 de febrero de 2019).

19. Citaremos a este respecto lo que señala la letra i) del artículo 2 del Real Decreto 1090/2015, de 4 de diciembre, por el que se regulan los ensayos clínicos con medicamentos, los Comités de Ética de la Investigación con medicamentos y el Registro Español de Estudios Clínicos define como «Ensayo clínico»: “*Un estudio clínico que cumpla cualquiera de las siguientes condiciones: (...) 3.º Se aplican procedimientos de diagnóstico o seguimiento a los sujetos de ensayo que van más allá de la práctica clínica habitual*”. A nuestro juicio, el uso de algoritmos destinados a determinar el tratamiento a aplicar a un paciente difícilmente podría entrar dentro de esta definición, por cuanto ni son procedimientos diagnósticos ni implican un seguimiento al paciente. Otra cosa sería, probablemente, el caso de los algoritmos que se enfocaran específicamente al diagnóstico.
20. US Food and Drug Administration (April 13, 2015), *Balancing Premarket and Postmarket Data Collection for Devices Subject to Premarket Approval Guidance for Industry and Food and Drug Administration Staff*, en: <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm393994.pdf>. Última consulta: 21 febrero de 2019.
21. Resolución del Parlamento Europeo, de 14 de marzo de 2017, Considerando 20.
22. Algunas sugerencias particularmente lúcidas figuran en el “discussion paper” del UK ICO, que señala que “Las técnicas de auditoría pueden utilizarse para identificar los factores que influyen en una decisión algorítmica. Los sistemas de visualización interactiva pueden ayudar a los individuos a entender por qué se hizo una recomendación concreta y darles control sobre futuras recomendaciones. Los comités de ética pueden utilizarse para ayudar a dar forma y mejorar la transparencia del desarrollo de algoritmos de aprendizaje automático. Debemos, en suma, utilizar sistemas de supervisión que combinen garanticen la transparencia de los algoritmos empleando a la vez un enfoque técnico y organizativo” (UK Information Commissioner’s Office 2017, 86)
23. Como escribe Cotino (2017, 139), “entiendo que este tipo de garantías frente a la discriminación van a tener que combinarse con el reconocimiento de fuertes facultades de acceso y conocimiento de los algoritmos y los grandes datos que se manejan por parte de sectores especializados, tanto públicos como de la sociedad civil (reguladores, académicos industria, asociaciones de consumidores, etc. (...)). Y dicha transparencia ha de venir acompañada asimismo con el reconocimiento de fuertes potestades de control a autoridades independientes respecto de los data brokers que son claves en el sector. Sin perjuicio del control social por la sociedad civil especializada, entiendo que son idóneas las autoridades independientes cercanas al sector de las nuevas tecnologías, como lo son las autoridades de protección de datos (y transparencia en EU, AEPD-ISMS, 2017,26 y ss.)”.
24. Al menos más allá de lo humanamente inevitable, ya que un mínimo sesgo se hallará siempre presente.
25. Como han señalado, entre otros, Boyd y Crawford (2011), los números no hablan por sí mismos; las afirmaciones de objetividad y precisión son engañosas dado que todos los investigadores son intérpretes de datos y siempre hay un proceso de “limpieza de datos” inherentemente subjetivo. En el mismo sentido, a la anterior idea se sumaría Surden (2017).
26. Como señala el Considerando M de la Resolución del Parlamento Europeo, de 14 de marzo de 2017, “debe hacerse una distinción entre cantidad y calidad de los datos a fin de facilitar la utilización eficaz de los macrodatos (algoritmos y otras herramientas analíticas); y que los datos y/o los procedimientos de baja calidad en los que se basan los procesos de toma de decisiones y las herramientas analíticas podrían dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales, el riesgo de utilización de los datos con fines discriminatorios o fraudulentos y la marginación del papel de los seres humanos en esos procesos, lo que puede traducirse en procedimientos deficientes de toma de decisiones con repercusiones negativas en las vidas y oportunidades de los ciudadanos, en particular los grupos marginalizados, así como generar un impacto negativo en las sociedades y empresas”.
27. Nos haremos eco aquí de lo señalado en la Consideración General número 31 de la Resolución del Parlamento Europeo, de 14 de marzo de 2017, que “Advierte de que, debido al efecto intrusivo de las decisiones y medidas adoptadas por las autoridades encargadas de la aplicación de la ley, también sobre la base del tratamiento de datos y la analítica de datos, en la vida de los ciudadanos y en sus derechos, se requiere la máxima prudencia para evitar la discriminación ilícita y la selección de determinadas personas o grupos de personas, definidos en relación con la raza, el color, el origen étnico o social, las características genéticas, la

lengua, la religión o creencia, las opiniones políticas o de otra índole, la propiedad, el nacimiento, la discapacidad, la edad, el género, la expresión o la identidad de género, la orientación sexual, la situación en materia de residencia, la salud o la pertenencia a una minoría nacional, que a menudo es objeto de elaboración de perfiles en función de la etnia o de una actuación policial más intensa, además de las personas que pueden definirse por características particulares; pide que se forme de manera adecuada a quienes recopilan los datos en primera línea y a los usuarios de las informaciones obtenidas del análisis de datos”.

- 28.** La capacidad de los algoritmos para inferir eficazmente datos no proporcionados, a menudo personales e incluso pertenecientes al tipo de los especialmente protegidos, a partir de otros que no tienen ese carácter resulta, de hecho, una amenaza latente para la distinción tradicional entre las diversas categorías de datos. Como señala el Considerando J de la Resolución del Parlamento Europeo, de 14 de marzo de 2017, con la aparición de los algoritmos inteligentes, “puede observarse la aparición de una difuminación entre datos personales y no personales derivada del uso de la analítica de macrodatos, que puede generar la creación de nuevos datos personales”. Más aún, parece del todo acertada su Consideración General 3, cuando señala que “puede inferirse información delicada sobre las personas de datos que no sean de carácter sensible, lo que desdibuja la línea entre la información delicada y la que no es de carácter sensible”. Piénsese, en este sentido, en la dificultad de diferenciar entre datos de salud y los que no deben ser considerados como tales. El Reglamento los define en su Considerando 35 como “Entre los datos personales relativos a la salud se deben incluir todos los datos relativos al estado de salud del interesado que dan información sobre su estado de salud física o mental pasado, presente o futuro. Se incluye la información sobre la persona física recogida con ocasión de su inscripción a efectos de asistencia sanitaria, o con ocasión de la prestación de tal asistencia, de conformidad con la Directiva 2011/24/UE del Parlamento Europeo y del Consejo (1); todo número, símbolo o dato asignado a una persona física que la identifique de manera unívoca a efectos sanitarios; la información obtenida de pruebas o exámenes de una parte del cuerpo o de una sustancia corporal, incluida la procedente de datos genéticos y muestras biológicas, y cualquier información relativa, a título de ejemplo, a una enfermedad, una discapacidad, el riesgo de padecer enfermedades, el historial médico, el tratamiento clínico o el estado fisiológico o biomédico del interesado, independientemente de su fuente, por ejemplo un médico u otro profesional sanitario, un hospital, un dispositivo médico, o una prueba diagnóstica *in vitro*”. Sin embargo, y como hemos indicado, un algoritmo puede utilizar datos que no tienen nada que ver con éstos para indicar el tratamiento más adecuado. En tal caso, ¿serían datos de salud todos los utilizados, aunque tal vez algunos de ellos (por ejemplo, el número de usos de una tarjeta de crédito) no parezcan tener nada que ver con la salud humana?
- 29.** Nótese que la distinción entre dato y dato deducido es sumamente significativa en el nuevo marco normativo. Como señala la Agencia Española de Protección de Datos, “el derecho a la portabilidad de datos (artículo 20) entendido como el derecho del afectado a obtener, en un formato estructurado, de uso común y lectura mecánica, la información que le concierna y haya facilitado a un responsable de tratamiento cuando esa información se trata por medios automatizados y sobre la base del consentimiento o para la ejecución de un contrato. En relación con este último derecho, recientemente el Grupo de Trabajo del Artículo 29 de la Directiva 95/46/CE ha adoptado directrices sobre la aplicación del derecho a la portabilidad. En este sentido, por ejemplo, el Grupo considera que el concepto de datos facilitados por el interesado incluye los datos proporcionados de manera activa por el interesado y los datos observados (datos de ubicación, búsqueda, ritmo cardíaco, etc) pero no incluye dentro de los datos sujetos al derecho a la portabilidad a los datos inferidos o deducidos que hayan sido creados por el responsable de tratamiento a partir de los datos proporcionados por el interesado (como pueden ser los resultados algorítmicos)”.
- 30.** Así, por ejemplo, ha alcanzado cierta notoriedad recientemente el caso de Tammy Dobbs, residente en Arkansas, EE. UU., que vio reducidas sus horas de atención sanitaria como consecuencia de la introducción de un mecanismo de decisión automatizado construido sobre un algoritmo, sin que mediase ninguna intervención humana en la determinación final. Véase: <https://www.theverge.com/2018/3/21/17144260/health-care-medicaid-algorithm-arkansas-cerebral-palsy>. Última consulta: 18 de febrero de 2019.
- 31.** Véase, al respecto: “Computer system could kill rather than cure, doctors warn”, en: <http://cphpost.dk/news/computer-system-could-kill-rather-than-cure-doctors-warn.html>. Última consulta: 18 de febrero de 2019.

- 32.** A este respecto, hay que aludir necesariamente al Reglamento Europeo de Protección de Datos, que ha eliminado radicalmente la posibilidad de implementar mecanismos automatizados de decisiones, esto es, el uso discriminatorio de la inteligencia artificial sin intervención alguna del componente humano. En este sentido, es obligatorio referirse a su artículo 22.1, que dice, textualmente, lo siguiente: “1. *Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar*”. Como cabe apreciar, su redacción ilegaliza cualquier forma de toma de decisiones en la que no intervenga de alguna forma el elemento humano, una prohibición que rige automáticamente, sin necesitar de una oposición explícita o implícita del afectado.
- 33.** A este respecto, merece la pena recordar lo dispuesto en el artículo 22(1) del Reglamento Europeo de Protección de Datos o el Considerando N de la Resolución del Parlamento Europeo de 14 de marzo de 2017, que dice que “la responsabilidad algorítmica y la transparencia deben implicar la aplicación de medidas técnicas y operativas que garanticen la transparencia, la no discriminación de la toma de decisiones automatizadas y el cálculo de probabilidades de comportamiento individual; que la transparencia debe facilitar a las personas información fiable sobre la lógica aplicada, el significado y las consecuencias previstas; y que debe incluir información sobre los datos utilizados para la formación en materia de análisis de macrodatos y permitir a las personas comprender y controlar las decisiones que los afectan”.
- 34.** Véase, a este respecto: Article 29 Data Protection Working Party, Opinion 3/2010 on the principle of accountability, Adopted on 13 July 2010, en: <http://www.dataprotection.ro/servlet/ViewDocument?id=654>. Última consulta: 18 de febrero de 2019.
- 35.** Así, por ejemplo, hay algoritmos, como el desarrollado por Genevès et al. (2017), que parecen capaces de predecir el riesgo de que los pacientes padezcan complicaciones hospitalarias utilizando datos provenientes de sus historiales y de la prescripción farmacológica. Obviamente, en tal caso no son necesarios datos personales adicionales.
- 36.** Una intervención destinada a extirpar la lesión puede resultar innecesaria y traumática, pero difícilmente causará la muerte del paciente. En cambio, la decisión de no intervenir, en caso de basarse en una estimación incorrecta, puede permitir al tumor producir una metástasis y acabar con la vida del paciente. Obviamente, esto sólo ocurrirá en un número muy escaso de individuos, pero un número escaso es mayor que cero.
- 37.** Nótese que hablamos sólo de la obtención de datos nuevos, pero no de la utilización de datos que el sistema sanitario ya posee y que desea emplear incluso sin el consentimiento del paciente. El régimen jurídico de esta cuestión se hallaría resuelto a nuestro juicio por lo dispuesto en el artículo 9.2.h) del REPD que dice que el tratamiento de datos personales genéticos, biométricos o relativos a la salud no está prohibido cuando el tratamiento es necesario “para fines (...) de diagnóstico médico, prestación de asistencia o tratamiento de tipo sanitario o social o gestión de los sistemas y servicios de asistencia sanitaria y social”.
- 38.** En el ordenamiento jurídico español, sin ir más lejos, esta obligación se refleja en lo dispuesto en el artículo 2.7 de la Ley de Autonomía del Paciente, que certifica la obligación que tiene todo paciente de facilitar los datos sobre su estado físico o sobre su salud de manera leal y verdadera, así como el de colaborar en su obtención, especialmente cuando sean necesarios por razones de interés público o con motivo de la asistencia sanitaria. De manera que, en principio, la existencia de este precepto sería razón suficiente como para negar el derecho a preservar la intimidad del sujeto