ARBOR Ciencia, Pensamiento y Cultura
Vol. 197-802, octubre-diciembre 2021, a628 | ISSN-L: 0210-1963
https://doi.org/10.3989/arbor.2021.802006

VALORACIONES ÉTICAS PARA UNA INTELIGENCIA ARTIFICIAL ADECUADA A LA PRIVACIDAD

ETHICAL ASSESSMENTS FOR A PRIVACY-FRIENDLY ARTIFICIAL INTELLIGENCE

Ricardo Morte Ferrer
Universidad de Granada
http://orcid.org/0000-0002-8002-9764
ricardo63@autistici.org

Cómo citar este artículo/Citation: Morte Ferrer, Ricardo (2021). Valoraciones éticas para una inteligencia artificial adecuada a la privacidad. Arbor, 197(802): a628. https://doi.org/10.3989/arbor.2021.802006

Recibido: 1 marzo 2021. Aceptado: 3 octubre 2021.
Publicado: 3 febrro 2022.

RESUMEN: Desde hace ya bastante tiempo existe una tendencia a afirmar que el derecho o las normas de diferente tipo no son adecuadas o aplicables para las diferentes nuevas tecnologías que van apareciendo de forma continua (computación en la nube, *big data*, Internet de las cosas, robots, inteligencia artificial...). Este trabajo intentará dar la vuelta a ese razonamiento y, centrándose en la inteligencia artificial, tratará de plantear criterios adecuados para que esa tecnología y muchas otras sean desarrolladas y aplicadas de forma adecuada a los derechos fundamentales en general y a la privacidad en particular. Dentro de esos criterios se analizará lo que debe aportar la explicabilidad de la tecnología mencionada (requisito introducido por diferentes grupos de expertos) y se profundizará en algunos conceptos adicionales como la posibilidad y la necesidad de controlar (hasta el punto de poder detener en cualquier momento un sistema de inteligencia artificial) y de auditar sistemas de inteligencia artificial. Para realizar la mencionada profundización se recurrirá a lo que se conoce como objetivos de protección (disponibilidad, confidencialidad, integridad, transparencia, no encadenabilidad y capacidad de intervenir) y a los principios de la ciberética, incluyendo una referencia especial a la sostenibilidad y a la resiliencia de este tipo de sistemas.

PALABRAS CLAVE: Inteligencia artificial, privacidad, derechos fundamentales, objetivos de protección.

ABSTRACT: For quite some time now, there has been a tendency to claim that the law or different types of standards are not adequate or applicable to the different new technologies that are continuously appearing (cloud computing, big data, the Internet of Things, robots, Artificial Intelligence, etc.). This paper will attempt to turn this reasoning on its head and, focusing on Artificial Intelligence, propose suitable criteria for this technology and many others to be developed and applied appropriately to fundamental rights, in general, and privacy in particular. Within these criteria, an analysis will be made of what the explainability of the aforementioned technology should provide (a requirement introduced by different groups of experts) and some additional concepts, such as the possibility/necessity of controlling (to the point of being able to stop an Artificial Intelligence system at any time) and auditing Artificial Intelligence systems, will be explored in greater depth. In order to carry out the aforementioned in-depth study, we will use what are known as protection goals (availability, confidentiality, integrity, transparency, unlinkability, and ability to intervene) and the principles of Cyberethics, including a special reference to the sustainability and resilience of these types of systems.

KEYWORDS: Artificial Intelligence, privacy, fundamental rights, protection goals.

1. INTRODUCCIÓN

En el resumen del presente trabajo se ha mencionado la tendencia existente a afirmar que el derecho y otras normas no son adecuadas o aplicables a las nuevas tecnologías, aunque hay que reconocer que en otros casos quienes eso piensan lo expresan de una forma más prudente y dicen que la normativa no es capaz de adaptarse con la velocidad necesaria a esas nuevas tecnologías. Aunque ese tema puede dar para muchas discusiones y se podría profundizar mucho, dicho de forma general quien se tiene que adaptar son las nuevas tecnologías, que deben cumplir la normativa vigente como lo debe hacer cualquier ciudadano.

En materia de privacidad (por utilizar la terminología más en boga, aunque la protección de datos o el derecho fundamental a la autodeterminación informativa podrían ser más adecuados[1]) el criterio a seguir se fijó en la Europa continental ya en los años setenta del siglo XX. En la que se considera como la primera ley reguladora de esta materia, el *Landesdatenschutzgesetz* del Land alemán de Hessen (1970)[2] se introdujo lo que se conoce como *Verbot mit Erlaubnisvorbehalt* (prohibición con excepción de autorización), que supone que el tratamiento de datos personales está prohibido salvo que se disponga de un mecanismo legal (consentimiento del sujeto afectado, contrato, ley, interés legítimo demostrable y documentado del responsable del tratamiento). Ahí podemos ver la valoración ética que se hizo desde un principio: el bien jurídico a proteger son los derechos fundamentales de los sujetos cuyos datos se pretende procesar. Conviene recordar que los derechos fundamentales son otorgados por el Estado a los ciudadanos para protegerse de la actividad del mismo Estado. Esto es especialmente importante porque las grandes empresas de las tecnologías de la información y de la comunicación pretenden hacer creer que el principio actúa a la inversa: autorización con excepción de prohibición.

A lo largo del presente trabajo se presentará lo que se conoce como objetivos de protección como criterios de control de cualquier tecnología, se revisarán diferentes aspectos del problema planteado por la inteligencia artificial y se recordarán algunos de los instrumentos ya existentes para que la implementación de esta tecnología se produzca de una forma respetuosa con los derechos fundamentales.

2. OBJETIVOS DE PROTECCIÓN

La doctrina de los objetivos de protección no es nueva, se desarrolló ya en el campo de la seguridad de la información en los años setenta del Siglo XX[3] centrándose en tres objetivos que, en inglés en el original, se conocen como CIA *(confidentiality, integrity, availability)*, lo cual no deja de resultar irónico en un trabajo que se centra en la privacidad. Posteriormente se han desarrollado otros centrados en la privacidad. A continuación se expondrán las características esenciales de estos objetivos.

Confidencialidad. Este objetivo de protección recoge como exigencia que nadie pueda acceder a los datos personales sin autorización. En ocasiones el acceso a los datos permite que el sujeto afectado sea identificado porque el contexto en el que los datos son almacenados permite sacar conclusiones sobre ese sujeto. Cuando nos referimos a personas no autorizadas, eso no significa que se trate necesariamente de terceros ajenos a la organización, que pueden actuar con intenciones criminales o de otro tipo, sino que puede tratarse también de empleados de servicios técnicos que para prestar esos servicios no precisan de acceso a los datos personales, o de personas activas en departamentos de la organización que no tienen ninguna relación con un determinado proceso o con el sujeto afectado.

Integridad. En este caso el objetivo de protección resalta como exigencia que los procesos y sistemas informáticos sean capaces de mantener las características que son esenciales para la realización de las funciones imprescindibles para alcanzar la finalidad establecida y, al mismo tiempo, que los datos tratados permanezcan indemnes, completos y actuales. Posibles efectos secundarios deben ser evitados o tenidos en cuenta y tratados. Este objetivo de protección reclama que entre las exigencias y la realidad haya una garantía suficiente, tanto en los detalles técnicos como en lo que afecta al tratamiento en general y su ajuste a las finalidades establecidas.

1 Aunque no parece necesario profundizar en esta diferenciación terminológica, sí puede valer la pena mencionar que el término privacidad está más relacionado con la concepción anglosajona del right to be let alone y de la esfera privada, mientras que en este trabajo se hará referencia a la concepción más extendida en la Europa continental.

2 https://www.rv.hessenrecht.hessen.de/bshe/document/jlr-DSGHErahmen Fecha de consulta: 3 octubre 2021.

3 https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nbsspecialpublication500-19.pdf Fecha de consulta: 3 octubre 2021.

ARBOR Vol. 197-802, octubre-diciembre 2021, a628 | ISSN-L: 0210-1963 https://doi.org/10.3989/arbor.2021.802006

Disponibilidad. Este objetivo refleja la exigencia de que los datos personales estén disponibles para ser utilizados de forma adecuada en el proceso previsto. Para ello deben ser accesibles para las personas correspondientes y se les deben poder aplicar los métodos indicados para su tratamiento. Eso incluye, entre otras cosas, que los métodos sean aplicables al formato en el que los datos están disponibles. La disponibilidad incluye que los datos sean localizables, que los sistemas implicados los puedan presentar de forma adecuada y que esa presentación sea semánticamente comprensible.

Antes de pasar a mencionar los nuevos objetivos de protección cabe recordar que los tres primeros han sido aceptados e implementados por los responsables de tratamiento por iniciativa propia, ya que, aunque pueden utilizarse también para proteger la privacidad en origen, están diseñados para proteger a las organizaciones frente a ataques tanto internos como externos.

Conviene reseñar que desde el punto de vista de la normativa de protección de datos, las organizaciones deben proteger sus procesos de posibles ataques, siempre que afecten a datos de carácter personal. Los objetivos de la protección de datos precisan, en comparación con los objetivos de protección de la seguridad informática, de un grado de comprensión más amplio, ya que la protección de datos debe incluir una perspectiva de protección adicional, al tener en cuenta los riesgos que las actividades de la organización en sí mismas pueden originar para el sujeto afectado, tanto si esas actividades están relacionadas con sus procesos de negocio o de administración como fuera de ellos. Desde el punto de vista metodológico eso significa que no solo una persona debe demostrar ante una organización que es de confianza, sino que la organización debe ser capaz de demostrar frente a una persona que es de confianza. Por ese motivo es preciso establecer objetivos de protección que garanticen la protección de los sujetos afectados frente a diferentes tipos de organizaciones, que describiremos a continuación.

No encadenabilidad. Este objetivo refleja la exigencia de que los datos solo sean tratados y valorados para la finalidad para la que fueron recogidos.

Transparencia. La transparencia requiere que, aunque en diferentes niveles, tanto el sujeto afectado como el responsable de los sistemas y posibles autoridades de control puedan reconocer qué datos y para qué finalidad han sido recogidos y tratados en un proceso, qué sistemas y procesos han sido utilizados, en qué dirección y para qué fines fluyen los datos y quién es el responsable legal de los datos y sistemas en las diferentes fases de un tratamiento de datos. La transferencia es imprescindible para el control y dirección de los datos, procesos y sistemas desde su inicio hasta su cancelación, y un requisito previo para que un tratamiento de datos sea legítimo y, en caso de necesidad, los sujetos afectados puedan otorgar su consentimiento.

La transparencia de un tratamiento de datos en su conjunto y de las partes implicadas puede permitir que especialmente los sujetos afectados y las autoridades de control puedan detectar posibles fallos y exigir que se lleven a cabo las modificaciones necesarias para suprimirlos.

Capacidad de intervenir. Este objetivo exige que el sujeto afectado pueda ejercer de forma efectiva sus derechos de acceso, rectificación, cancelación y oposición (ARCO) en cualquier momento, y que el responsable esté obligado a tomar las medidas necesarias para hacer efectivos esos derechos. Para alcanzar este objetivo debe ser posible modificar el tratamiento de datos en cualquier momento y en cualquiera de sus fases, desde la recogida de los datos hasta su cancelación.

3. RIESGOS RELACIONADOS CON LA INTELIGENCIA ARTIFICIAL

Normalmente cuando hablamos de problemas de seguridad en materia de información y comunicación, lo primero que consideramos es la confidencialidad, cosa que no extraña si pensamos en privacidad. Sin embargo, dada la amplitud de la implementación de sistemas de inteligencia artificial, cabe recordar que los problemas en materia de integridad y disponibilidad son como mínimo igual de importantes. Baste para ello recordar que uno de los ámbitos en los que parece que los sistemas que nos ocupan se van a implementar más rápido es en el de la salud y más concretamente en el de la radiología[4], motivo por el que la integridad y disponibilidad ganan una relevancia mayor. Aunque no parece necesario profundizar en los aspectos técnicos del problema, qui-

4 El autor participó en la consulta previa a la elaboración de un informe al respecto https://www.berlin-university-alliance.de/commitments/knowledge-exchange/kira-bericht.pdf Fecha de consulta: 3 octubre 2021.

zás un ejemplo que apareció en prensa[5] sirva para reflejar el problema: por medio de un ataque informático un grupo de científicos introdujo imágenes falsas de nódulos cancerosos en resultados de tomografía computarizada para llamar la atención sobre el problema. También existe un artículo científico al respecto (Mirsky *et al*., 2019).

Un campo distinto es el uso de sistemas de inteligencia artificial para procedimientos de reconocimiento facial. El *New York Times* publicó un artículo[6] sobre la tecnología utilizada por la empresa Clearview AI cuyo título ya avisa del problema al que hace referencia: «The Secretive Company That Might End Privacy as We Know It». Por exponerlo de forma muy resumida, esa empresa desarrolló un software que podía extraer imágenes de prácticamente cualquier base de datos disponible en Internet (incluyendo Facebook y Youtube, por mencionar dos casos significativos) y que fue adquirido por numerosas agencias gubernamentales. Todo eso se ve agravado por varias circunstancias: parece que el sistema solo alcanza un 75% de exactitud, no ha proporcionado ninguna auditoría externa y no informa sobre las fuentes y procedimientos por los que obtiene las imágenes.

Otro aspecto esencial a tener en cuenta hace referencia a la sostenibilidad de los sistemas de inteligencia artificial. La mayoría de los estudios se centran en los problemas que origina el transporte de mercancías, mientras que lo que sucede en Internet o en las TIC goza de un halo de limpieza que se hace difícil justificar. A día de hoy el tráfico de datos gasta tanta energía como todo el tráfico aéreo[7] (conviene recordar que la mayoría de las aplicaciones disponibles en los llamados teléfonos inteligentes están basadas en servidores y no en el terminal) y está previsto que en el 2040 el World Wide Web gaste tanta energía como consumió todo el tráfico mundial, no solo el aéreo, en el 2011[8]. Por poner otro ejemplo, la producción de teléfonos inteligentes podría generar en ese momento 125 mega toneladas de CO^{29}. Parece evidente que se hará necesario despedirse de un sistema económico orientado al crecimiento sin fin y que la continua búsqueda de una *bala de plata* tecnológica que solucione esos problemas de momento no va en la dirección adecuada, ya que las presuntas soluciones están aumentando los problemas o creando otros nuevos, basta pensar en lo que tecnologías como el *blockchain* y el bitcoin consumen a nivel energético. Desgraciadamente a día de hoy parece que el número de investigadores que se ocupan al mismo tiempo de las nuevas tecnologías y de su impacto ecológico es realmente reducido, con consecuencias que pueden ser desastrosas. Revisemos por ejemplo el problema del transporte de personas y su presunta solución: los mal llamados vehículos autónomos (en realidad teledirigidos, ya que dependen de servidores). A día de hoy nadie ha explicado a fondo cómo van a circular al mismo tiempo vehículos autónomos y otros con un conductor humano, y tampoco como van a convivir con otros vehículos de menor calidad como las bicicletas. En realidad parece que se trata de una solución inexistente y que habría que buscar una alternativa. A primera vista lo más fácil sería construir ciudades adecuadas para peatones y bicicletas, junto con sistemas públicos de transporte de calidad y asequibles. Incluso la ficción parece ofrecer una solución mejor: Frank Schätzing en su novela *Limit*[10] incluye un sistema de vehículos individuales que circulan por raíles (y que personalmente creo mucho más viable que los vehículos autónomos) y, aunque no tiene nada que ver con el tema actual, un sistema de viaje a la luna por medio de un ascensor.

Por último, ya que no pretendemos hacer un listado completo de los riesgos relacionados con la inteligencia artificial, mencionaremos un punto mucho más importante de lo que puede parecer en un principio. Para entender su importancia conviene recordar que uno de los objetivos de los profetas de la inteligencia artificial es conseguir conectar directamente las mentes humanas con los sistemas, que deberán ser capaces de leer (¿o quizás fuera mejor decir interpretar?) nuestros pensamientos. Hacemos referencia a los sistemas de reconocimiento de emociones, sobre los que recientemente se ha publicado un informe[11] sobre su evolución en China.

También parece adecuado mencionar la propuesta de reglamento del Parlamento Europeo y del Conse-

5 https://www.washingtonpost.com/technology/2019/04/03/hospital-viruses-fake-cancerous-nodes-ct-scans-created-by-malware-trick-radiologists/ Fecha de consulta: 3 octubre 2021.

6 https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html Fecha de consulta: 3 octubre 2021.

7 https://www.fr.de/wirtschaft/mit-einem-hurrikan-kann-nicht-verhandeln-12272668.html Fecha de consulta: 3 octubre 2021.

8 https://brighterworld.mcmaster.ca/articles/how-smartphones-are-heating-up-the-planet/ Fecha de consulta: 3 octubre 2021.

9 https://www.welt.de/wissenschaft/article13391627/Wie-das-Internet-zum-Klimakiller-wird.html Fecha de consulta: 3 octubre 2021.

10 https://de.wikipedia.org/wiki/Limit_(Roman) Fecha de consulta: 3 octubre 2021.

11 https://www.article19.org/emotion-recognition-technology-report/ Fecha de consulta: 3 octubre 2021.

jo, por la que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión[12], que plantea la prohibición de determinados sistemas de inteligencia artificial y establece algunos requisitos como la necesidad de implementar un sistema de gestión de riesgos, la necesidad de transparencia en la información con los usuarios y la vigilancia humana. Por otra parte, también hay que resaltar que se trata de una propuesta y que habrá que ver cómo queda la redacción final, así como la dificultad que supone aplicar una regulación de este tipo a sistemas ya implementados.

4. INSTRUMENTOS PARA UN DESARROLLO ADECUADO DE LA INTELIGENCIA ARTIFICIAL

A continuación mencionaremos algunos de los ejemplos de instrumentos que pueden ayudar, ya en la fase de diseño, a conseguir un desarrollo controlado de la tecnología que nos ocupa.

4.1 Evaluación de impacto sobre la privacidad

Empezaré por comentar que el término alemán *Datenschutzfolgenabschätzung*, evaluación de las consecuencias para la protección de datos (traducción propia), parece más adecuado, ya que incluye la evaluación del impacto y la de las consecuencias de un tratamiento de datos personales considerado de riesgo para los sujetos afectados.

Esta es una de las novedades esenciales del reglamento general de protección de datos (RGPD) y está regulada en su artículo 35[13]. A continuación se incluirán los aspectos de esa regulación que parecen esenciales para el tema que nos ocupa:

El número 1. Cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del mismo, una evaluación del impacto de las operaciones en la protección de datos personales. Una única evaluación podrá abordar una serie de operaciones de tratamiento semejantes que entrañen altos riesgos similares.

El número 7. La evaluación deberá incluir como mínimo: a) una descripción sistemática de las operaciones de tratamiento previstas y de los fines del tratamiento, inclusive, cuando proceda, el interés legítimo perseguido por el responsable del tratamiento; b) una evaluación de la necesidad y la proporcionalidad de las operaciones de tratamiento con respecto a su finalidad; c) una evaluación de los riesgos para los derechos y libertades de los interesados a que se refiere el apartado 1, y d) las medidas previstas para afrontar los riesgos, incluidas garantías, medidas de seguridad y mecanismos que garanticen la protección de datos personales, y a demostrar la conformidad con el presente reglamento, teniendo en cuenta los derechos e intereses legítimos de los interesados y de otras personas afectadas.

En el primer punto queda claro que el objetivo de control no es la nueva tecnología en sí misma, sino que el núcleo de esta figura es la protección de los derechos fundamentales de los sujetos afectados por el tratamiento de sus datos personales.

El siguiente apartado recoge el contenido mínimo de una evaluación de este tipo y reitera su finalidad.

Es esencial recordar que esta evaluación se debe realizar siempre antes de iniciar el tratamiento de datos personales *reales* en el proceso sometido a evaluación y que esa evaluación forma parte del fundamento legal del mismo (recordemos que sin una base legal adecuada el tratamiento de datos personales está prohibido) y que si no se realiza o se realiza de forma incorrecta el tratamiento será ilegal desde el primer momento. También es un punto esencial de esta evaluación el que al acabarla debe programarse ya cuándo se repetirá y que se deben implementar los controles necesarios para detectar cuándo será necesario repetirla, incluso antes de la fecha prevista, por ejemplo debido a un cambio en los riesgos a tener en cuenta, bien sea por modificaciones en la organización, en las tecnologías utilizadas o en el contexto social afectado.

Dado que en este trabajo se están estudiando los riesgos que pueden plantear los sistemas de inteligencia artificial, parece adecuado mencionar el documento directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento «entraña probablemente un alto riesgo» a efectos del reglamento (UE) 2016/679 adoptado por el grupo de trabajo del artículo 29 (hoy comité europeo de protección de datos) en 2017[14]

12 https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021PC0206&from=EN Fecha de consulta: 3 octubre 2021.
13 https://www.privacy-regulation.eu/es/35.htm Fecha de consulta: 3 octubre 2021.
14 https://www.aepd.es/sites/default/files/2019-09/wp248rev01-es.pdf Fecha de consulta: 3 octubre 2021.

4.2 Evaluación del impacto ético

Aquí se aporta una lista con los puntos a tener en cuenta para el modelo propuesto por el autor del presente trabajo en un artículo previo (Morte Ferrer, 2020) en el que se ampliaban los principios de ciberética planteados por Romero Muñoz (2017), y en el que se puede revisar la fundamentación del modelo.

1. *Explicabilidad*. Este principio ya se vió incluido en las *AI Ethics Guidelines*[15] del grupo de expertos de la Comisión Europea, como ya se ha mencionado anteriormente en este trabajo. Pero aquí se quiere resaltar la importancia de este principio no solo en lo que afecta a la inteligencia artificial, sino para cualquiera de las nuevas tecnologías aplicadas al ámbito de la salud. El significado de este principio radica en que toda organización que pretenda implementar una nueva tecnología en el ámbito que nos ocupa debe alcanzar el máximo grado de transparencia posible sobre el procedimiento a implementar. La regla esencial debería ser que lo que no es explicable no debería ser implementado, si bien se pueden aceptar excepciones debidamente documentadas y que deberían ser revisadas de forma continuada para garantizar el grado de transparencia adecuado.
2. *Autonomía y libertad del usuario* de las nuevas tecnologías. Esto supone que el usuario debe poder conocer las herramientas tecnológicas a fin de poder decidir de forma libre sobre si las quiere usar o no. Este principio tendría como consecuencia que el software utilizado en el campo de la salud debería ser libre o como mínimo de código abierto para poder controlarlo de forma adecuada. El software privativo no es controlable.
3. *Beneficiencia*. Este principio trata de garantizar la obligación de actuar en beneficio de otros, promoviendo sus legítimos intereses y suprimiendo prejuicios. No parece preciso profundizar en este punto, especialmente cuando estamos estudiando temas de salud.
4. *No maleficencia*. Este principio obliga a abstenerse intencionadamente de realizar actos que puedan causar daño o perjudicar a otros. Se reitera lo expuesto en el punto anterior.
5. *Justicia*. Este principio conduce e tratar a cada uno como corresponda, con la finalidad de disminuir las situaciones de desigualdad. Se debe interpretar como la capacidad de garantizar que la nueva tecnología va a funcionar de forma justa, sin crear desigualdades. De nuevo la utilización del software libre parece esencial para el cumplimiento de este principio.
6. *Sostenibilidad*. Siguiendo este principio, hay que comprobar y analizar el impacto tecnológico sobre la contaminación del suelo, la atmósfera, y el sistema de reciclado de materiales.
7. *Precaución*. Paralelamente al principio de no maleficencia, el de precaución respalda la adopción de medidas de ciberseguridad y de protección ante la sospecha de que ciertas tecnologías puedan crear riesgos en el futuro. En lo que afecta a la ciberseguridad, se debe garantizar la integridad, la disponibilidad y la confidencialidad de los sistemas a implementar. En lo que afecta a la disponibilidad cabe resaltar que se debe garantizar la redundancia de sistemas para prever, por ejemplo, qué sucederá cuando un sistema inteligente falle. Como ya se ha comentado anteriormente, si en el campo de la radiología se reduce la tarea de los médicos a administrar los resultados emitidos por un dispositivo de inteligencia artificial, se hará difícil poder garantizar que esos médicos serán capaces de realizar diagnósticos correctos evaluando ellos mismos las imágenes.
8. *Privacidad*. Para que se cumpla este principio, el usuario debe conocer los mecanismos de privacidad en red por su seguridad y anonimato, así como los sistemas de privacidad de hardware y software. Aquí se debe garantizar de nuevo la integridad, disponibilidad y confidencialidad de los sistemas, pero poniendo el énfasis en los pacientes. Siempre hay que recordar que la seguridad se centra en las organizaciones, mientras que la privacidad lo hace en las personas, en nuestro caso pacientes. Además hay que garantizar la transparencia, la capacidad de intervenir y la no encadenabilidad.
9. *Democracia*. Para garantizar el cumplimiento del principio de autonomía, se debe promover

15 Ethics guidelines for trustworthy AI https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html Fecha de consulta: 3 octubre 2021.

ARBOR Vol. 197-802, octubre-diciembre 2021, a628 | ISSN-L: 0210-1963 https://doi.org/10.3989/arbor.2021.802006

en los órganos institucionales una defensa de los derechos digitales como derechos humanos, así como la ciberseguridad en las infraestructuras sanitarias. Se debe garantizar que los nuevos sistemas han sido diseñados e implementados asegurando el cumplimiento de esos derechos.

10. *Controlabilidad*. La organización que plantea la implementación de una nueva tecnología debe poder garanizar que es capaz de controlarla. Por explicarlo de forma gráfica, debe existir un botón rojo que pueda detener el sistema en cualquier momento en caso de necesidad. Este problema se ha planteado en varias ocasiones en los estudios sobre inteligencia artificial, en casos en los que el resultado obtenido es el esperado pero no se puede justificar cómo se ha obtenido. En el campo de la salud este aspecto es especialmente importante.

11. *Auditabilidad*. Este principio está estrechamente relacionado con el anterior, especialmente en el segundo aspecto comentado. Es necesario que se puedan documentar y justificar todas las medidas aplicadas según la nueva tecnología implementada.

4.3 Auditorías algorítmicas

Uno de los problemas planteados por los sistemas de inteligencia artificial es que en muchas ocasiones constituyen lo que se conoce como *black box*, término que refleja la falta de transparencia y de explicabilidad de estos sistemas. Una auditoría de este tipo da la sensación de ser un intento de control *a posteriori* de los sistemas que nos ocupan, figura que va contra el modelo legal vigente y contra el modelo propuesto por el autor en el artículo anteriormente mencionado. De todas formas puede ser un modelo útil para controlar y reducir los problemas planteados por sistemas diseñados e implementados sin las evaluaciones de impacto adecuadas.

A día de hoy parece que el modelo más adecuado es el planteado por la empresa Eticas Consulting[16], disponible en la red previo registro de una dirección de correo electrónico.

De acuerdo con la información disponible en la página web de la empresa:

> «La Guía de Auditoría Algorítmica tiene tres objetivos principales:
>
> Proteger los derechos fundamentales relativos a la privacidad y la protección de datos personales;
>
> Aportar claridad a las leyes aplicables a los sistemas algorítmicos;
>
> Ofrecer una metodología para controlar que estas tecnologías son diseñadas, desarrolladas y utilizadas no solo de acuerdo a la Ley, sino de forma socialmente justa y responsable».

A continuación aportaremos lo mencionado en la guía que nos ocupa sobre recomendaciones relativas al cumplimiento ético y legal:

> «De forma general, el cumplimiento de los derechos fundamentales a la privacidad y la protección de datos personales debe ser respetado y, en la medida de lo posible, promovido, tanto en los procesos de diseño, desarrollo e implementación de un algoritmo, como durante el proceso de auditoría. Esto debe ser así también para todos aquellos derechos que se puedan ver afectados en el caso concreto de un algoritmo».

Si bien hay que reconocer el trabajo realizado en esta guía, la formulación de los requisitos parece demasiado débil para poder afrontar el problema que plantean los sistemas de inteligencia artificial en los aspectos que nos ocupan en el presente trabajo.

4.4 Documentos de la Agencia Española de Protección de Datos

La Agencia Española de Protección de Datos ha publicado documentos relacionados con los sistemas que nos ocupan.

En primer lugar, apareció el documento titulado «Adecuación al RGPD de tratamientos que incorporan inteligencia artificial. Una introducción»[17] en el que la Agencia declara no pretender realizar un repaso exhaustivo de lo regulado en el RGPD, sino simplemente dar apoyo a las organizaciones que utilizan sistemas de inteligencia artificial. De nuevo hay que reconocer el trabajo realizado, que también de nuevo no parece ser el adecuado para afrontar realmente el problema.

Posteriormente se publicó el documento sobre: «Requisitos para Auditorías de Tratamientos que incluyan IA»[18], que también da la sensación de ser un in-

16 https://www.eticasconsulting.com/eticas-consulting-guia-de-auditoria-algoritmica-para-desarrollar-algoritmos-justos-y-eficaces/ Fecha de consulta: 3 octubre 2021.

17 https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf Fecha de consulta: 3 octubre 2021.

18 https://www.aepd.es/sites/default/files/2021-01/requisitos-auditorias-tratamientos-incluyan-ia.pdf Fecha de consulta: 3 octubre 2021.

tento de reducir y controlar en la medida de lo posible los problemas generados por sistemas de inteligencia artificial, sin ir al origen de los posibles problemas. Un documento de este tipo debería estar basado en la evaluación del impacto sobre la privacidad que cualquier sistema de este tipo debería presentar y sin la cual ningún dato debería ser tratado.

5. CINCO CONSEJOS PARA EMPRESAS QUE QUIERAN UTILIZAR INTELIGENCIA ARTIFICIAL

Este apartado nace de un artículo aparecido recientemente[19] y que parece hacer necesario expresar también cinco consejos en una dirección diferente:

- Nunca inicies un tratamiento de datos personales en un sistema de inteligencia artificial sin estar totalmente seguro de disponer de una base legal adecuada y sin haber realizado una evaluación del impacto sobre la privacidad antes de iniciar el tratamiento.
- Si realmente pretendes utilizar la inteligencia artificial para mejorar el futuro de la humanidad no te preocupes de los temas de propiedad intelectual, utiliza sistemas basados en software libre, y si tienes ánimo de lucro, intenta generar tus ingresos por medio de servicios relacionados con el sistema y no por medio del sistema en sí mismo.
- Recurre siempre a sistemas de software libre, que permiten auditar todo el proceso y que luchan contra el problema del *black box*, permitiendo controlar el desarrollo y en caso de necesidad parar el sistema (recordemos el objetivo de protección de la capacidad de intervenir).
- Crea tu modelo de negocio en torno a los servicios y no en torno al desarrollo del sistema, de esa forma aumentará la probabilidad de conseguir un sistema de inteligencia artificial que ayude a más gente.
- Nunca recurras al modelo de patentes. Probablemente un modelo de cooperación inteligente pueda generar más beneficios que un sistema de protección equivocada del conocimiento propio (aunque sea una anotación al margen, este punto está relacionado con la situación pandémica actual y con los problemas que las patentes están generando).

6. CONCLUSIONES

Parece relativamente claro que el problema planteado por la inteligencia artificial está relacionado con el sistema económico en el que ha aparecido. Es necesario pasar a un sistema de crecimiento reducido, o quizás incluso de decrecimiento. Este punto no está en el núcleo del tema del presente trabajo, pero tiene una relación directa con el mismo. Recordemos el tema de la sostenibilidad y del impacto del tráfico de datos. Siguiendo el modelo actual, puede parecer interesante o incluso aconsejable que muchos teléfonos inteligentes sean capaces de utilizar sistemas de inteligencia artificial. Sin embargo, probablemente lo interesante sería utilizar la inteligencia artificial tanto como sea necesario y tan poco como sea posible, para intentar optimizar (no maximizar) sus efectos y reducir sus consecuencias negativas. Lo mismo cabe decir de sistemas como el 5G, por mencionar un ejemplo no directamente relacionado con la inteligencia artificial pero sí con una nueva tecnología que plantea problemas muy similares, que se están desarrollando básicamente porque pueden generar beneficios, pero sin realizar las evaluaciones de impacto que serían aconsejables y que seguramente habrían tenido como consecuencia que se promoviera el desarrollo de la fibra óptica y no el del 5G[20].

La aplicación de los objetivos de protección mencionados en el presente trabajo desde la fase de diseño de cualquier sistema de inteligencia artificial permitiría controlar el desarrollo del mismo y poder despedirnos del problema del *black box,* ya que un sistema que no se ajustara a los requisitos planteados por esos objetivos no debería implementarse en ningún caso. En el texto se han mencionado los avances en la implementación de sistemas de inteligencia artificial en el ámbito de la salud y más concretamente en el de la radiología. Si revisamos este caso desde el punto de vista de los objetivos de protección parece que debería surgir prácticamente de forma inmediata una pregunta relacionada con el objetivo de protección de la disponibilidad (ya se ha comentado antes algo sobre la integridad): ¿qué sucedería si el sistema de inteligencia artificial no estuviera disponible, bien por una avería o bien por un fallo del sistema eléctrico? A primera vista puede parecer que no es un problema grave, pero si anteriormente se ha convertido a los radiólogos en meros gestores de los resultados del sis-

19 https://www.jdsupra.com/legalnews/five-tips-for-life-sciences-companies-3589104/ Fecha de consulta: 3 octubre 2021.

20 https://www.eff.org/deeplinks/2019/10/why-fiber-vastly-superior-cable-and-5g Fecha de consulta: 3 octubre 2021.

tema, el problema es mucho mayor de lo que puede parecer en un principio.

El tema de las evaluaciones de impacto debería jugar un papel esencial para conseguir una mejora en el desarrollo de sistemas de inteligencia artificial y de otras nuevas tecnologías que todavía están por aparecer. Aplicando un principio del ámbito analógico, «es más fácil construir bien que tener que reformar *a posteriori*», parece evidente que esas evaluaciones son el instrumento ideal para controlar los programas de desarrollo e implementación de los sistemas que nos ocupan. Muchos de los problemas que vivimos en la actualidad en materia de seguridad relacionados con Internet tienen su raíz en un desarrollo incontrolado que nos ha llevado a una situación que se refleja en el título del libro de Schneier (2018): *Click here to kill everybody*. Lo mismo cabe decir en lo referente a las evaluaciones sobre el impacto ético que, por poner un ejemplo, podrían permitir detectar problemas de diseño en los puntos mencionados en el modelo incluido en el presente trabajo. Aprovechando este punto para hacer una referencia a la pandemia del COVID-19, que actualmente afecta al mundo, podemos recordar las diferentes aplicaciones de localización basadas en la tecnología *bluetooth* que se han desarrollado en diferentes países. Los inventores de esta tecnología advirtieron ya en mayo de 2020[21] que no era la adecuada para la finalidad perseguida, y la mayoría de las aplicaciones basadas en esa tecnología han fracasado. De hecho, al fin y al cabo, las medidas que realmente han funcionado son bastante antiguas, principalmente ha sido el confinamiento y la distancia social las que han mostrado una mayor efectividad, unas medidas que ya se conocían en la Edad Media. Esta afirmación puede parecer provocadora, pero es muy real. Había otras opciones tecnológicas menos problemáticas y menos agresivas que las aplicaciones con tecnología *bluetooth*, como listas basadas en un código QR con un período de almacenamiento limitado, pero de nuevo la búsqueda de la bala de plata llevó al camino y a la solución equivocada.

El problema de la sostenibilidad también parece bastante claro. Se hace necesario llevar a cabo una priorización de los sistemas que realmente son imprescindibles y los que no lo son. De nuevo el mecanismo de las evaluaciones de impacto sería el adecuado para llevar a cabo esa priorización. Los datos aportados sobre el consumo de energía del tratamiento y tráfico de datos dejan claro que se hace necesario optimizar la utilización de los recursos energéticos, porque de seguir por el camino actual parece inevitable que se confirme lo que algunos autores mencionan:

> «A los problemas de sostener la red y la potencia eléctrica globalmente, se sumarán los de su infraestructura (de Internet), redes de fibra óptica, satélites, ordenadores. Se pasará de la era de Internet a la de la radio, en el mejor de los escenarios tecnológicos posibles. Primero se perderá la neutralidad en la red, un proceso que ya ha comenzado. Después, a pesar de su centralidad estratégica, fallará la financiación pública y privada (sobre todo vía publicidad). En la siguiente etapa, la red se irá empequeñeciendo, restringiéndose el acceso a quien no se pueda pagar la infraestructura, la conexión o, simplemente, no tenga enganche a la red eléctrica. Finalmente caerá» (Fernández Durán y González Reyes, 2014, II: 292).

O como lo ha expresado Riechmann (2016): «Internet caerá. Pero si cayese ahora, el desastre sería menos dañino que si lo hace dentro de 10 años, y dentro de 10 años menos que si ocurre dentro de 20 años».

Seguramente lo mismo cabe decir respecto de la utilización del software libre para sistemas de inteligencia artificial y de comunicaciones. Sería el camino adecuado para optimizar, que no maximizar, recursos y para abandonar un sistema basado en la competitividad por sí misma para pasar a uno basado en la cooperación en el acceso del mayor número de ciudadanos a los recursos disponibles. De nuevo la pandemia del COVID-19 nos muestra que la cooperación y la optimización de recursos es el camino a seguir. Como ya se ha mencionado anteriormente, el sistema de patentes y de la venta al mejor postor (es otro campo, pero el mismo problema) puede complicar la salida de la pandemia, mientras que una distribución equitativa de las vacunas y una inversión adecuada en terapias, no solo en vacunas, para el tratamiento de la enfermedad facilitarían esa salida de la pandemia y reducirían la desigualdad[22].

21 https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/ Fecha de consulta: 3 octubre 2021.

22 https://ctxt.es/es/20210201/Firmas/35100/antivirales-vacunas-investigacion-coronavirus-pandemia-manuel-calleja.htm Fecha de consulta: 3 octubre 2021.

AGRADECIMIENTOS

Este artículo ha sido elaborado en el marco de sendas colaboraciones con los proyectos INBOTS (European Union's Horizon 2020 research and innovation programme under grant agreement No. 780073) y EXTEND (European Union's Horizon 2020 research and innovation programme under grant agreement No. 779982).

BIBLIOGRAFÍA

Fernández Durán, Ramón; González Reyes, Luis (2014). *En la espiral de la energía*. Madrid: Libros en acción.

Mirsky, Yisroel; Mahler, Tom; Shelef, Ilan, and Elovici, Yuval (2019). *CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning*. Disponible en https://arxiv.org/pdf/1901.03597.pdf [Fecha de consulta: 3 octubre 2021].

Morte Ferrer, Ricardo (2020). Reflexiones sobre las evaluaciones del impacto. Una propuesta para un modelo de Evaluación del Impacto Ético en el ámbito de salud. *Dilemata. Revista Internacional de Éticas Aplicadas*, 32: 71-82. Disponible en https://www.dilemata.net/revista/index.php/dilemata/article/view/412000351 [Fecha de consulta: 3 octubre 2021]

Riechmann, Jorge (2016). ¿Derrotó el smartphone al movimiento ecologista? *Para una crítica del mesianismo tecnológico*. Madrid: Los Libros de la Catarata.

Romero Muñoz, Javier (2017). CiberÉtica como ética aplicada: una introducción. *Dilemata. Revista Internacional de Éticas Aplicadas*, 24: 45-63. Disponible en https://www.dilemata.net/revista/index.php/dilemata/article/view/412000100 [Fecha de consulta: 3 octubre 2021].

Schneier, Bruce (2018). *Click Here to Kill Everybody. Security and Survival in a Hyper-connected World*. W. W. Norton & Company.