Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial.
Una introducción

Febrero de 2020

RESUMEN EJECUTIVO

Este documento tiene como objetivo ser una primera aproximación para la adecuación al Reglamento (UE) 2016/679, General de Protección de Datos (RGPD) de productos y servicios que incluyan componentes de Inteligencia Artificial. Bajo la etiqueta “Inteligencia Artificial”, o IA, subyacen muchos tipos de soluciones que emplean distintas técnicas, destacando entre ellas las basadas en aprendizaje automático. La IA ha se ha convertido en un componente más de los tratamientos de datos realizados por los responsables y que, en muchos casos, aparece en forma de soluciones desarrolladas por terceros. La Inteligencia Artificial genera muchas dudas entre los usuarios, investigadores, especialistas, autoridades y la industria con relación a aspectos de cumplimiento normativo, respeto a los derechos de los interesados y seguridad jurídica de todos los intervinientes. Estas dudas representan una dificultad para el correcto desarrollo tecnológico.

El presente documento no pretende realizar un repaso exhaustivo a lo establecido en el RGPD, pero sí abordar las dudas planteadas en el marco de protección de datos de carácter personal y señalar los aspectos más relevantes en la relación IA-RGPD que deben ser tenidos en cuenta desde el diseño y en la implementación de tratamientos que incluyan IA. En el documento se prestará especial atención, entre otros, a la legitimación para el tratamiento, la información y transparencia, el ejercicio de derechos, las decisiones automatizadas, la exactitud, la minimización de datos, la evaluación de impacto y el análisis de la proporcionalidad del tratamiento. El texto está dirigido a responsables que incluyan componentes de IA en sus tratamientos, así como desarrolladores, encargados u otros, que den soporte a dichos tratamientos.

Palabras clave: Inteligencia artificial, IA, RGPD, LOPDGDD, derechos, privacidad, ética, protección de datos, diseño, evaluación de impacto, aprendizaje automático, machine learning, ML, decisiones automatizadas, minimización, transparencia, exactitud, sesgo.

ÍNDICE

I. INTRODUCCIÓN AL MARCO IA Y PROTECCIÓN DE DATOS

El término Inteligencia Artificial, o IA, fue usado por primera vez en 1956 por John McCarthy para referirse a "la ciencia y la ingeniería de crear máquinas inteligentes, especialmente programas de computación inteligentes". El Grupo de Alto Nivel en Inteligencia Artificial (AI – HLEG) que ha creado la Comisión Europea para desarrollar la Estrategia Europea en Inteligencia Artificial lo aplica a "*sistemas que manifiestan un comportamiento inteligente, al ser capaces de analizar el entorno y realizar acciones, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos*"[1]. Otras definiciones apuntan que la IA es la capacidad de una máquina de actuar como lo hace una mente humana, considerando sus aspectos de creatividad y capacidad de realizar análisis e inferencias[2] a partir de información compleja, e incluso incompleta.

En función del alcance y el ámbito de aplicación de la inteligencia artificial se diferencian tres categorías diferentes de IA: las inteligencias artificiales fuertes, generales y débiles. La IA general podría resolver cualquier tarea intelectual resoluble por un ser humano; la IA fuerte o superinteligencia iría más allá de las capacidades humanas. Pero el tipo de IA que ha disparado la aplicación práctica de esta disciplina es la que se conoce como IA-débil (*AI-weak*) y que, en contraste con la IA fuerte y general, se caracteriza por desarrollar soluciones capaces de resolver un problema concreto y acotado[3]. La aplicación de este tipo de sistemas es extensa: desde los videojuegos a sistemas de defensa, pasando por entorno sanitario, control industrial, robótica, buscadores de Internet, tratamiento de lenguaje natural, marketing, asistentes personales, recursos humanos, optimización de servicios públicos, gestión energética, medioambiente y cualquier otra actividad que nos podamos imaginar[4]. El ámbito de aplicación de las soluciones IA se extiende a todos los sectores, cada uno de ellos con casuísticas específicas y con la necesidad de cumplir tanto con una normativa general como con la normativa sectorial.

La aplicación de la Inteligencia Artificial despierta dudas entre usuarios, investigadores, especialistas, autoridades y la industria con relación a aspectos de cumplimiento normativo, respeto a los derechos de los interesados y seguridad jurídica de todos los intervinientes. Estas dudas pueden ser un obstáculo para el correcto desarrollo tecnológico, por lo que es necesario desarrollar guías y ayudas que aborden y resuelvan las dificultades encontradas. En este documento nos centraremos en tratar la adecuación al RGPD de tratamientos que incorporan componentes de IA-débil.

A. TÉCNICAS DE IA

Existen diversas formas de aproximarse a una solución de IA: mediante redes neuronales, sistemas basados en reglas, lógica borrosa, aprendizaje automático, sistemas expertos, sistemas adaptativos, algoritmos genéticos, sistemas multiagente, etc., términos que se solapan unos con otros[5]. Lo que pretenden todas estas técnicas es conseguir modelos para tratar sistemas complejos para los que no se sabe cómo, o no es posible tratar con algoritmos secuenciales, por su dificultad para modelar comportamientos regidos por múltiples variables. Entre estas variables se establecen relaciones no lineales, o que no es posible aproximar a métodos lineales, y que incluso pueden variar con el tiempo.

[1] Artificial Intelligence for Europe, Comisión Europea
[2] RAE. Inferencia: acción y efecto de inferir. Inferir: Deducir algo o sacarlo como conclusión de otra cosa.
[3] La inferencia en ese caso puede intentar resolver un problema de clasificación, como la identificación de individuos sospechosos; o de un problema de "clustering", como recomendar artículo a un cliente basándonos en compras anteriores; o un problema de regresión (estimación de un valor), como detectar la intenciones de voto de un colectivo.
[4] En AI Index 2018 Annual Report http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf se puede consultar un informe sobre la extensión y mercado de los componentes de IA
[5] Se puede crear un sistema experto modelando el conocimiento de una serie de expertos mediante reglas en lógica borrosa.

Una de las ramas de la IA con más éxito en aplicaciones comerciales es el Aprendizaje Automático o Machine Learning (ML). El ML diseña modelos predictivos que construyen por sí mismos la relación entre las variables a estudiar mediante el análisis de un conjunto inicial de datos, la identificación de patrones y el establecimiento de criterios de clasificación. Una vez fijados los criterios, al introducir un nuevo conjunto de datos el componente IA es capaz de realizar una inferencia. El aprendizaje automático está, relacionado con las técnicas de minería de datos, optimización y *big data*[6]. A su vez, existen distintos tipos de aprendizaje automático como el supervisado[7], no supervisado[8], de refuerzo[9] y sus variantes, que emplean distintas técnicas. Así mismo se encuentran especializaciones del ML como el Deep Learning o Aprendizaje Profundo[10], y diferentes modelos de aprendizaje, como el centralizado, el descentralizado o el federado. Un sistema con un componente de IA podemos decir que es adaptativo cuando el modelo de inferencia se ajusta dinámicamente en función de cada nuevo conjunto de datos de entrada, refinando las relaciones ya establecidas.

B. TRATAMIENTOS CON COMPONENTES DE IA

Puede haber tratamientos que incluyen componentes de IA que manejan datos de personas físicas, como en un modelo de perfilado de marketing o electoral, o puede haber tratamientos en los que no aparezcan datos de carácter personal, como podría suceder en un modelo de predicción meteorológico que recoge datos de estaciones geográficamente distribuidas.

Un tratamiento que tome decisiones automatizadas usando la inteligencia artificial puede afectar a personas físicas, como por ejemplo un sistema de autenticación de usuarios, o puede no afectar a personas, como un sistema de control industrial. En el que caso de que se tomen decisiones que afectan a las personas, estas decisiones pueden ser relativas a la interacción de la persona en su contexto social, como el acceso a un contrato o servicio, o relativas a la personalización de dicho servicio, como podría ser la personalización en los mandos de un coche o la programación de un televisor. Las decisiones pueden hacer predicciones sobre la evolución del sujeto, realizar una evaluación sobre el estado actual de éste, o bien decidir la ejecución de un conjunto de acciones[11].

A su vez, en la toma de decisiones, la IA puede adoptar dos roles:

- El de una ayuda para el proceso de decisión, proporcionando una inferencia o perfil sobre un sujeto o una situación, para que un ser humano tome la decisión final.
- El de toma y ejecución de la decisión.

En cualquiera de los casos anteriores, el componente IA no va a estar aislado, sino que va a formar parte de un tratamiento más amplio. Además, tendrá una implementación física en un sistema en el que se encontrarán otras aplicaciones, existirán comunicaciones de datos, interfaces de usuario, y otros elementos. La sinergia inherente en las aplicaciones en donde se encuentran componentes IA puede relacionar estos tratamientos con Big Data,

[6] Aunque con un conjunto de datos de alta capacidad predictiva es posible realizar ML con "*small data*".
[7] El sistema se entrena con un conjunto de ejemplos en los que los resultados de salida son conocidos los algoritmos trabajan con datos "etiquetados" intentado encontrar una función que, dadas las variables de entrada, les asigne la etiqueta de salida adecuada.
[8] Pretende la extracción de información significativa, sin la referencia de variables de salida conocidas, y mediante la exploración de la estructura de dichos datos sin etiquetar. Parten de datos no etiquetados en los que no existe información de clasificación ni de evento dependiente continuo. La labor del algoritmo es la de encontrar la estructura interna de los datos.
[9] Existe intervención humana durante el proceso de aprendizaje premiando o penalizando las decisiones parciales.
[10] Técnica de ML que emplea múltiples capas de tratamiento no lineal de la información, cada capa adaptada a capturar mediante aprendiza una característica determinada, y organizadas en una jerarquía desde un nivel de abstracción más bajo a uno más alto.
[11] Por ejemplo, en el ámbito sanitario, prever la evolución de un paciente, realizar un diagnóstico automático tras exploración o prescribir un determinado tratamiento.

Internet de las Cosas (IoT), 5G/sistemas móviles, Edge Computing[12] y Computación en la Nube. En definitiva, intervienen múltiples elementos técnicos, pero también humanos, de forma directa o colateral, que hay que tener en cuenta para determinar las implicaciones que tiene la utilización de un tratamiento que incorpora una solución IA.

Finalmente, hay que tener en cuenta que no todo sistema que toma una decisión automatizada es IA, no toda IA es Machine Learning, ni todo lo que se publicita como IA es realmente IA[13], sino que dicha etiqueta puede ser un recurso de marketing o se puede emplear para implementar otro tipo de estrategias de negocio.

C. PROTECCIÓN DE DATOS Y LA DIMENSIÓN ÉTICA

La protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental. El artículo 8, apartado 1, de la Carta de los Derechos Fundamentales de la Unión Europea y el artículo 16, apartado 1, del Tratado de Funcionamiento de la Unión Europea establecen que toda persona tiene derecho a la protección de los datos de carácter personal que le conciernan[14]. El derecho fundamental a la protección de datos está desarrollado en un marco normativo que actualmente comprende el Reglamento 619/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD), y se complementa en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD) además de toda la normativa sectorial publicada antes y después de la entrada en vigor del RGPD.

La perspectiva ética de la IA, como una parte de la "ética digital", es uno de los aspectos que más inquietud despierta[15]. La ética de la IA persigue proteger valores como la dignidad, la libertad, la democracia, la igualdad, la autonomía del individuo y la justicia frente al gobierno de un razonamiento mecánico. La Comisión Europea trabaja en la definición de una Inteligencia Artificial confiable, y establece que para ello, ha de cumplir con siete requisitos clave: acción y supervisión humanas, solidez técnica y seguridad, gestión de la privacidad y los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas.

Estos requisitos deben ser evaluados a lo largo de todo el ciclo de vida de un sistema de IA de forma continua. Es preciso ser vigilantes tanto sobre la legitimidad ética de los tratamientos como de los efectos inesperados de estos. Asimismo, debe considerarse el posible impacto colateral de dichos tratamientos en un entorno social, más allá de las limitaciones concebidas inicialmente de propósito, de duración en el tiempo y de extensión.

Es decir, hay que analizar la solución IA *per se*, pero también en el marco del tratamiento en el que se integra, y las relaciones de dicho tratamiento con el entorno en varios aspectos:

- En el aspecto cultural, con su escala de valores.
- En el contexto en el que se despliega el servicio, con sus requisitos de calidad.
- En los aspectos que se derivan de la interconexión masiva de componentes en la sociedad de la información.

Un aspecto crítico de los sistemas de IA es el de la posible existencia de sesgos. Un sesgo ("*bias*" en inglés) es una desviación inadecuada en el proceso de inferencia. Los sesgos son particularmente graves cuando, por ejemplo, derivan en discriminaciones de un

[12] Tratamiento que se realiza en la ubicación física del dispositivo del usuario, de la fuente de datos, o cerca de ellas.

[13] El timo de la Inteligencia Artificial: el 40% de las empresas lo usan como reclamo para financiarse https://www.elmundo.es/tecnologia/2019/03/12/5c829c0d21efa0760a8b45fa.html . DANIEL J. OLLERO, 12de marzo de 2019

[14] Considerando 1 del RGPD

[15] Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Cathy O'Neil, Broadway Books 2016

grupo en favor de otro[16]. Este problema no es particular de los sistemas de IA, sino que es general a cualquier proceso de toma de decisión, ya sea humano o automático.

Pero existe otro tipo de sesgo que puede ser aún más preocupante y que es el del sesgo en la interpretación de los resultados de la IA. Este sesgo humano consiste en aceptar, sin espíritu crítico, los resultados de una IA como ciertos e inamovibles, asumiendo un "principio de autoridad" derivado de las expectativas creadas por dichos sistemas.

La extensión y profundidad de los valores éticos de las personas depende en gran medida del entorno cultural en el que se desarrollan como personas. La privacidad y la intimidad de los individuos forma una parte importante de los principios éticos y, de igual manera, la percepción que cada sociedad puede tener de dichos valores puede llegar a ser muy distinta[17], aunque se han ido acordando internacionalmente unos principios aceptados de forma global[18].

D. DEFINICIONES EN EL RGPD

En el RGPD se establecen definiciones y conceptos marco, fundamentalmente en su artículo 4 y, sin dejar de remitir a dicho artículo, algunos de ellos los desarrollamos a continuación que son especialmente interesantes con relación a la IA:

Objeto del RGPD

El artículo 1 del RGPD estable el objeto de la norma como la protección de derechos y libertades fundamentales de las personas físicas, en particular con relación a la protección de datos de carácter personal[19].

Dato personal

El concepto de dato personal se define en el artículo 4.1 del RGPD como *"toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona;"*

Esta definición se desarrolla con más profundidad en el Dictamen 4/2007 sobre el concepto de datos personales del Grupo del Artículo 29 (WP29).

---

[16] *"systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate" (*Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330-347).

[17] En algunos países el acceso a la información fiscal es uno de los aspectos más privados de la persona, mientras que en otros es información pública: "Noruega, el país donde nadie puede esconder su salario" https://www.bbc.com/mundo/noticias-internacional-40691744

[18] Artículo 12 de la Declaración Universal de los Derechos del Hombre. Convenio para la Protección de los Derechos Humanos y Libertades Fundamentales, 1950. Pacto de los Derechos Civiles y Políticos, 1966. Resolución 509 de la Asamblea del Consejo de Europa sobre los derechos humanos y los nuevos logros científicos y técnicos. Recomendación de la OCDE sobre la circulación internacional de datos personales para la protección de la intimidad, 1980 y actualizada en 2002. Convenio 108 del Consejo de Europa, para la protección de las personas con respecto al tratamiento automatizado de sus datos de carácter personal, 1981. Resolución de 14 de enero de 1990 de la Asamblea General de las Naciones Unidas, relativa a los principios rectores para la reglamentación de los ficheros computerizados de datos personales …

[19] Artículo 1 Objeto. 1. El presente Reglamento establece las normas relativas a la protección de las personas físicas en lo que respecta al tratamiento de los datos personales y las normas relativas a la libre circulación de tales datos. 2. El presente Reglamento protege los derechos y libertades fundamentales de las personas físicas y, en particular, su derecho a la protección de los datos personales. …

Seudonimización y anonimizacion

La seudonimización se define en el artículo 4.5 del RGPD como "*el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable*".

La anonimización es el proceso que permite eliminar o reducir al mínimo los riesgos de reidentificación de un individuo a partir de sus datos personales eliminando toda referencia directa o indirecta a su identidad, pero manteniendo la veracidad de los resultados del tratamiento de los mismos. Es decir, además de evitar la identificación de las personas, los datos anonimizados deben garantizar que cualquier operación o tratamiento que pueda ser realizado con posterioridad a la anonimización no conlleva una distorsión de los datos reales[20].

Ambas definiciones se completan en el Dictamen 05/2014 sobre técnicas de anonimización del WP29 y en las Orientaciones y garantías en los procesos de anonimización de datos personales publicada por la AEPD.

Categorías especiales de datos

Las categorías especiales de datos se establecen en el artículo 9 del RGPD como aquellos datos que "*que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos[21], datos biométricos[22] dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud[23] o datos relativos a la vida sexual o la orientación sexual de una persona física*".

El RGPD establece en dicho artículo la prohibición genérica a su tratamiento, prohibición que se amplía en el también artículo 9 de la LOPDGDD, salvo que se den una serie de circunstancias detalladas en los apartados 2, 3 y 4 del artículo 9 del RGPD y los apartados 1 y 2 del artículo 9 de la LOPDGDD.

Tratamiento

El artículo 4.2 del RGPD define tratamiento como cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción.

[20] "Orientaciones y garantías en los procesos de anonimización de datos personales" AEPD 2016

[21] Artículo 4.13: datos personales relativos a las características genéticas heredadas o adquiridas de una persona física que proporcionen una información única sobre la fisiología o la salud de esa persona, obtenidos en particular del análisis de una muestra biológica de tal persona

[22] Artículo 4.14: datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos

[23] Artículo 4.15: datos personales relativos a la salud física o mental de una persona física, incluida la prestación de servicios de atención sanitaria, que revelen información sobre su estado de salud

La elaboración de perfiles (ver definición a continuación) y la toma de una decisión sobre una persona física son tratamientos según los Considerandos 24[24] y 72[25] del RGPD.

Perfilado

El artículo 4.4 del RGPD define perfilado como una forma de tratamiento de datos personales que permite inferir más información acerca de una persona física, evaluando, analizando o prediciendo aspectos personales[26].

Un tratamiento que implique la elaboración de perfiles se caracteriza por tres elementos[27]:

- Debe ser una forma automatizada de tratamiento, incluyendo aquellos tratamientos que tienen participación parcialmente humana.
- Debe llevarse a cabo respecto a datos personales;
- Y el objetivo de la elaboración de perfiles debe ser evaluar aspectos personales sobre una persona física.

Decisiones automatizadas

Las decisiones basadas únicamente en el tratamiento automatizado representan la capacidad de tomar decisiones por medios tecnológicos sin la participación del ser humano[28]. El RGPD, en sus Considerandos 71 y 71 así como en su artículo 22, limita y establece derechos con relación a que los sujetos de los datos no sean sometidos a decisiones exclusivamente automatizadas[29] que tengan efectos jurídicos o que afecten significativamente al interesado. La elaboración de perfiles de forma automática se incluye en este marco de decisiones automatizadas. El Grupo del Artículo 29 ha analizado las implicaciones de este derecho en las Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679 del WP29.

No obstante, ha de tenerse en cuenta que las decisiones automatizadas pueden llevarse a cabo con o sin elaboración de perfiles y la elaboración de perfiles puede darse sin realizar decisiones automatizadas.

Usuarios de la solución IA

Tomando como punto de vista de protección de datos, los usuarios del tratamiento que utilizan IA podrían clasificarse de la siguiente forma:

- Entidades que emplean dicha IA sobre datos de interesados (empleados, clientes u otros) como podría ser una empresa que emplea IA para determinar políticas de marketing sobre las preferencias de sus clientes.

[24] Considerando 24 del RGPD: "… el potencial uso posterior de técnicas de tratamiento de datos personales que consistan en la elaboración de un perfil de una persona física con el fin, en particular, de adoptar decisiones sobre él o de analizar o predecir sus preferencias personales, comportamientos y actitudes."

[25] Considerando 72: La elaboración de perfiles está sujeta a las normas del presente Reglamento que rigen el tratamiento de datos personales, como los fundamentos jurídicos del tratamiento o los principios de la protección de datos. El Comité Europeo de Protección de Datos establecido por el presente Reglamento (en lo sucesivo, el «Comité») debe tener la posibilidad de formular orientaciones en este contexto.

[26] Artículo 4.4 «elaboración de perfiles»: toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física.

[27] Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679. Grupo del Artículo 29

[28] Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679

[29] Para ser pueda considerarse que existe participación humana, la supervisión de la decisión ha de ser realizada persona autorizada y competente para modificar la decisión, y ha de realizar una acción significativa y no simbólica.

- Personas físicas que adquieren un producto o un servicio que incluye un componente de IA con el objeto de tratar sus propios datos personales, como podría ser un individuo que compra una pulsera de actividad que planifica su entrenamiento.

Responsable

La figura de responsable de un tratamiento se define en el RGPD, en el artículo 4.7[30] del RGPD como la persona que determine los fines y medios del tratamiento, y el ámbito de sus obligaciones, que se enmarcan en el artículo 24, incluye, entre otras, la de "*aplicar medidas técnicas y organizativas apropiadas a fin de garantizar y poder demostrar que el tratamiento es conforme con el presente Reglamento*".

Corresponsable

El RGPD introduce la figura del corresponsable del tratamiento en el artículo 26, como aquellos responsables (dos o más) que determinen conjuntamente los objetivos y los medios del tratamiento. Los corresponsables determinarán de modo transparente y de mutuo acuerdo sus responsabilidades respectivas en el cumplimiento de las obligaciones impuestas por el RGPD.

Encargado

El encargado del tratamiento es aquella persona que trate datos personales por cuenta del responsable, tal como se establece en el artículo 4.8[31] del RGPD, y el ámbito de sus obligaciones, que se enmarcan en el artículo 28, contempla, entre otras, que su relación con el responsable *"se regirá por un contrato u otro acto jurídico con arreglo al Derecho de la Unión o de los Estados miembros, que vincule al encargado respecto del responsable".*

Excepción doméstica

En el artículo 2.c se establece, y se desarrolla en el Considerando 18, que el RGPD no aplica cuando el tratamiento lo realice una persona física en el curso de una actividad exclusivamente personal o doméstica.

Estas actividades son, por ejemplo, "*la correspondencia y la llevanza de un repertorio de direcciones, o la actividad en las redes sociales y la actividad en línea realizada en el contexto de las citadas actividades"* y, en general, aquellas *"sin conexión alguna con una actividad profesional o comercial".*

Sin embargo, ha de tenerse presente que la excepción doméstica no aplica a los responsables o encargados del tratamiento que proporcionen los medios para tratar datos personales relacionados con tales actividades.

E. CICLO DE VIDA DE UN SOLUCIÓN IA

Una vez definidos algunos conceptos marco, se puede iniciar un análisis de los tratamientos que incluyen una solución o utilizan un componente IA como parte de estos.

[30] Artículo 4.7 «responsable del tratamiento» o «responsable»: la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros.

[31] Artículo 4.8 «encargado del tratamiento» o «encargado»: la persona física o jurídica, autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento.

Una solución IA será un elemento de proceso de datos que se incluirá en una o más fases de un tratamiento. En unos casos, el componente IA se desarrollará específicamente para dicho tratamiento, en otros muchos casos, dicho componente será desarrollado por terceros distintos del responsable[32]. El componente IA no estará aislado y se integrará en un tratamiento específico junto a otros componentes como: la recogida de datos, sistemas de archivos, módulos de seguridad, interfases de usuario, y otros. Es más, una vez desarrollado, un componente IA podría ser integrado en tratamientos de distintos responsables[33].

El ciclo de vida de un sistema IA[34], desde su génesis a su descarte, pasará a través de distintas etapas comunes a todos los desarrollos tecnológicos. No obstante, en función de la tecnología de IA que se implemente, podría tener algunos matices o particularidades. Estas etapas son:

- Concepción y análisis, en la que se fijan los requisitos funcionales y no funcionales de la solución IA. Estos vendrán fijados por objetivos de negocio derivados del tratamiento en donde se incorporará o del mercado donde se pretende comercializar el componente. Incluirá los planes de proyecto, las restricciones normativas, etc.
- Desarrollo: incluyendo etapas de investigación, prototipado, diseño, pruebas, entrenamiento y validación. No todas las etapas estarán siempre presentes y su existencia quedará supeditado a la solución de IA concreta adoptada. Por ejemplo, la etapa de entrenamiento sí estará presente en componentes IA basados en aprendizaje automático (ML).
- Explotación: esta etapa comprende la ejecución de distintas acciones, y algunas de ellas se ejecutarán en paralelo: integración, producción, despliegue, inferencia, decisión, mantenimiento y evolución[35].
- Retirada final del tratamiento/componente.

No hay que olvidar que, para el propósito de este documento, nos hemos centrado en el componente IA, pero a la hora de realizar un análisis del tratamiento, el componente IA y el resto de los elementos que conforman el tratamiento se han de estudiar como un todo. Además, en función del tipo de aplicación, algunas de las etapas anteriores podrían estar solapadas. Por ejemplo, la validación se podría solapar durante las etapas de desarrollo y explotación, o la etapa de evolución podría desarrollarse de manera simultánea a la etapa de inferencia.

F. TRATAMIENTOS DE DATOS PERSONALES USANDO IA

Adoptando una visión extensiva de los posibles tratamientos en una solución IA[36], se pueden encontrar datos personales en las siguientes etapas del ciclo de vida:

- **Entrenamiento**: si es un modelo IA basado, por ejemplo, en técnicas de ML, se podrían utilizar datos personales en el desarrollo de este. En otras ocasiones, como es el caso en el que se entrene un modelo IA mediante la captura de

[32] Como es, por ejemplo, el desarrollo de chat-bots o plataformas conversacionales de atención al cliente.
[33] Amazon Web Services ofrece módulos IA para incorporar en los tratamientos, y como se lee en su página "No machine learning required" https://aws.amazon.com/es/machine-learning/ai-services/.
[34] No hay que confundir las etapas del ciclo de vida de un componente IA, de un componente de un tratamiento, con la fases en las que se puede dividir un tratamiento para su análisis. La Guía Práctica para Evaluaciones de Impacto en la Protección de Datos de la AEPD, expone una división de los tratamientos en fases desde el punto de los datos en: captura de datos, clasificación/almacenamiento, cesión de datos a terceros y destrucción de los datos. Esta división es genérica, una primera aproximación para analizar un tratamiento en el momento de la explotación del mismo, y cada tratamiento tiene que adaptar esta división a sus condiciones específicas. Por ejemplo, la etapa de desarrollo de un componente IA basado en ML es un tratamiento con sus distintas fases que se deberán analizar.
[35] Mantenimiento evolutivo
[36] Como se ha señalado anteriormente, no todos ellos se encontrarán en la implementación de algunas soluciones IA

conocimiento de un experto, podría considerarse a priori que no existe un tratamiento de datos de carácter personal.

En el caso de que el entrenamiento trate datos de carácter personal, este es un tratamiento en sí mismo. En su máxima expresión podría incluir las siguientes actividades: definición, búsqueda y obtención del conjunto de datos de interés[37], preprocesamiento de la información (tratamiento de datos no-estructurados, limpieza, balanceo, selección, transformación), *splitting* o partición del conjunto de datos para verificación, e información de trazabilidad y de auditoría.

- **Validación[38]**: en esta operación se podría realizar un tratamiento de datos personales cuando se utilicen datos que corresponden a la situación real del tratamiento, para determinar la bondad del modelo de forma experimental. El conjunto de datos puede ser distinto de aquellos utilizados en la etapa de entrenamiento (si es que ésta existe y trata datos personales) y podría incluso ser realizada por un tercero para la auditoría o certificación del modelo.
- **Despliegue**: en el caso que la solución IA sea un componente, un módulo que se distribuye a terceros para incluir en sus tratamientos, se puede considerar que hay una comunicación de datos personales cuando la solución IA incluya datos de personales o exista una forma de obtenerlos. Por ejemplo, algunas soluciones IA, como las Máquinas de Soporte Vectorial (SVM[39]) podrían contener dentro de la lógica del modelo ejemplos de los datos de entrenamiento. En otros casos, se podrían encontrar patrones en el modelo que identifican a un individuo singular.
- **Explotación**: en las distintas actividades de explotación de la solución IA es posible encontrar tratamientos de datos personales:
    - **Inferencia**: cuando se usen datos del interesado para obtener un resultado, cuando se usen datos de terceros con el mismo propósito o cuando datos e inferencias del interesado se almacenan. Si el propio interesado dispone de la IA como un componente de su propiedad, aplicaría la excepción doméstica.
    - **Decisión**: como se ha visto anteriormente, la decisión sobre un interesado es un tratamiento de datos personales.
    - **Evolución**: en la solución IA se podrían usar los datos y resultados de los interesados para refinar el modelo de IA. Cuando nos encontramos que esa evolución se realiza en el componente adquirido por el propio interesado, de forma aislada y autónoma, aplicaría la excepción doméstica. Pero si se envían a terceros, tendríamos una comunicación de datos, un posible tratamiento de almacenamiento, tratamiento para modificar el modelo, o incluso nuevas comunicaciones si esos datos se incorporan al modelo y este es accesible a otros terceros.
- **Retirada**: la retirada del servicio puede tener dos extensiones distintas: el componente IA se retira por obsoleto en todos los tratamientos en los que se implemente, o un usuario concreto decide no utilizar el componente IA. Ese usuario puede ser una entidad o una persona física y puede tener efectos en la

[37] Forma parte de la actividad de minería de datos.

[38] La validación se puede realizar o complementar por otros métodos, como ejemplo, métodos analíticos. La aproximación de este texto se enfoca al uso de datos personales.

[39] Las máquinas de soporte vectorial, máquinas de vectores de soporte o máquinas de vector soporte (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje supervisado. Estos métodos están relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento etiquetados en clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases, más cercanos al que se llama vector soporte.

supresión local, centralizada o distribuida de datos, así como sobre la portabilidad del servicio.

Antes de continuar, hay que recordar que no todas las soluciones IA tratan datos personales en alguna de las etapas de su ciclo de vida, ni toman decisiones basadas únicamente en tratamientos automatizados que afectan a personas físicas[40]. Algunos ejemplos de soluciones IA sin datos personales podrían ser los sistemas de control de calidad de productos industriales, o aquellos sistemas de toma de decisiones sobre la compra y venta de productos financieros.

Si un componente IA realiza el tratamiento de datos personales, elabora perfiles sobre una persona física o si toma decisiones sobre la misma, tendrá que someterse al RGPD. En caso contrario, no será necesario. En muchos casos no es sencillo determinar si durante una etapa del ciclo de vida de un sistema basado en IA se tratan o no datos personales.

Durante el ciclo de vida de una solución IA pueden haberse usado datos personales de alguna forma, por ejemplo, en la etapa de desarrollo. En ese caso, dicha etapa constituye un tratamiento y está sujeta al cumplimiento del RGPD. En etapas posteriores del ciclo de vida de la solución IA, por ejemplo, cuando se integra en un tratamiento, hay que evaluar si se tratan datos personales para determinar si el tratamiento está sujeto al cumplimiento del RGPD, al menos con relación a la solución IA. Si se considera que no se tratan datos personales, por que estos se han eliminado o anonimizado, hay que demostrar que estos procesos han sido realmente efectivos y evaluar cuál es el riesgo de reidentificación.

Si queremos descartar que se tratan datos personales en etapas posteriores de su ciclo de vida, por ejemplo, cuando se integra la solución IA en un tratamiento, y que, por tanto, este tratamiento no está sujeto al cumplimiento del RGPD, hay que demostrar que la eliminación o anonimización de los datos personales es realmente efectiva y evaluar cuál es el posible riesgo de reidentificación que existe.

En la etapa de explotación del sistema que incorpora la solución IA nos podemos encontrar varias situaciones:

- El que ostenta la propiedad del sistema IA da acceso al mismo a los interesados sometidos al tratamiento de IA. Es, por ejemplo, el caso de la evaluación psicológica de los usuarios de Facebook que realiza la propia red social en determinados países[41].
- El que ostenta la propiedad del sistema IA decide traspasar los derechos de uso de dicha IA como un componente a un tercero, tratándose como un módulo *off-the-shelve*, desvinculándose de la explotación del sistema[42]. Un ejemplo podría ser un sistema IA de asistente a la conducción que un fabricante de coches adquiere como componente e incorpora en sus modelos[43].
- Una entidad decide los fines de un tratamiento, y contrata a quien ostenta la propiedad del sistema IA para que ejecute una de las fases de dicho tratamiento. De esta forma, el que ostenta la propiedad de la IA es quien trata de forma efectiva los datos personales bajo el encargo y las instrucciones de un responsable[44].

[40] La utilización de soluciones de IA para la toma de decisiones en el entorno financiero puede tener graves consecuencias jurídicas https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune
[41] https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/
[42] Aunque puede que no se desvincule de la evolución del sistema.
[43] https://igniteoutsourcing.com/automotive/artificial-intelligence-in-automotive-industry/
[44] https://es.wordpress.org/plugins/tags/captcha/

G. EVALUACIÓN DE LAS SOLUCIONES IA

Tanto el modelo de IA incluido en un tratamiento, como el tratamiento en sí, ha de tener el propósito de dar respuesta a una necesidad real de la empresa y la industria. Estamos hablando de soluciones que trascienden el ámbito experimental y se van a someter a las leyes del mercado, un mercado regulado y que está obligado a cumplir con unas normas y estándares de calidad[45].

Con relación a su capacidad de dar cumplimiento a los requisitos del tratamiento, hay ciertos parámetros comunes a cualquier solución técnica que deberán quedar especificadas, como, por ejemplo:

- Precisión, exactitud o medidas de error requeridos por el tratamiento[46].
- Requisitos de calidad en los datos de entrada al componente IA.
- Precisión, exactitud o medidas de error efectivas de la solución IA en función de la métrica adecuada para medir la bondad de esta[47].
- Convergencia del modelo, cuando nos encontremos con entrenamiento y soluciones adaptativas.
- Consistencia entre los resultados del proceso de inferencia.
- Predictibilidad del algoritmo[48] [49]
- Y cualquier otro parámetro de evaluación del componente IA.

Una solución técnica que no tenga respuesta a estas preguntas de una forma acreditable, no se podría considerar basada en una tecnología madura, sino en una tecnología sin capacidad de cumplir con los requisitos básicos de "accountability", transparencia y legalidad. Además, de esas respuestas se derivarán requisitos esenciales desde el punto de vista de protección de datos como es, por ejemplo, la aplicación del principio de minimización.

H. BREVE RESUMEN DE OBLIGACIONES QUE ESTABLECEN EL RGPD

El propósito de este documento no es reproducir el contenido del RGPD, para lo cual se remite al texto de la norma. De forma breve, el RGPD se desarrolla en 6 principios establecidos en el Capítulo II:

1. Licitud, lealtad y transparencia
2. Limitación de la finalidad (especificación del propósito)
3. Minimización de datos
4. Exactitud
5. Limitación del plazo de conservación
6. Integridad y confidencialidad

En el mismo capítulo se establecen las condiciones para que un tratamiento de datos personales sea legítimo.

---

[45] What's your ML Test Score? A rubric for ML production systems, Eric Breck, 2018, Google Inc. https://www.eecs.tufts.edu/~dsculley/papers/ml_test_score.pdf

[46] Estas medidas dependerán del tipo de modelo de IA utilizado, ya sea de clasificación, regresión u otros. Precisión se refiere a la dispersión del conjunto de valores obtenidos en las mediciones repetidas de una magnitud. Una medida común de la variabilidad es la desviación estándar de las mediciones. En modelos de clasificación la precisión se puede evaluar como la proporción de instancias correctamente clasificadas o con curvas ROC. La exactitud está relacionada con el sesgo de una estimación y se refiere a cuán cerca del valor real se encuentra el valor medido. La exactitud se puede modelar con los parámetros de error total, error medio, error absoluto medio o error cuadrático medio

[47] Esta precisión y exactitud podría exceder los requisitos del tratamiento, o no.

[48] La predictibilidad permite realizar una declaración precisa comportamiento del componente IA en determinadas condiciones especificadas. Aunque el "Problema de la Parada" afirma que no existe una manera automática computable de saber si todos los programas posibles terminan, no se niega que existan métodos de prueba para el análisis de componentes concretos.

[49] No hay que confundir la impredecibilidad factual de un algoritmo debido a la falta de análisis de este, el desconocimiento de los estados internos o la carencia de unos procedimientos de prueba exhaustivos con la aleatoriedad.

En el Capítulo III se establecen el conjunto de derechos que asisten a los sujetos de los datos, la obligación de garantizarlos y de implementar mecanismos efectivos para el ejercicio de estos, en particular los derechos de transparencia, información, acceso, rectificación, supresión, limitación, oposición, portabilidad y, uno más que tiene gran importancia en determinadas aplicaciones de IA como son los derechos que tienen los ciudadanos con relación a la toma de decisiones automatizadas.

En el Capítulo IV se establece el modelo de responsabilidad y cumplimiento establecido basado en la “accountability”, o responsabilidad proactiva, y cuyos elementos rectores son:

- La identificación de una responsabilidad en el tratamiento.
- El análisis del riesgo para los derechos y libertades.
- El estudio de la necesidad y la proporcionalidad de las operaciones de tratamiento con respecto a su finalidad.
- El despliegue de medidas para la gestión del riesgo, medidas de privacidad por defecto y desde diseño, medidas de seguridad, de gestión de incidentes, etc.

Para terminar este breve repaso, en el Capítulo V se establecen las condiciones para la ejecución de transferencias de datos personales a terceros países u organizaciones internacionales.

II. ROLES, RELACIONES Y RESPONSABLES

Uno de los aspectos claves del RGPD, y capital para determinar la correcta aplicación de las políticas de accountability y transparencia, es que estén claramente identificadas las responsabilidades en el tratamiento.

En las distintas etapas del ciclo de vida de un componente IA será responsable del tratamiento de datos personales aquella persona física, jurídica, autoridad pública u otro que tome la decisión de realizar el tratamiento de datos personales, como se ha definido anteriormente[50]. Por lo tanto, distintas responsabilidades implicarán distintas obligaciones en el marco del tratamiento. Es posible que dicho responsable contrate a terceras partes para realizar, en su nombre y bajo sus instrucciones, diferentes tareas. Dichos terceros tendrán el carácter de encargados de tratamiento siempre y cuando todo tratamiento de datos personales lo realicen bajo las instrucciones de ese responsable. Cualquier otro tratamiento adicional sobre dichos datos que puedan llegar a realizar para sus propios fines los convertirá en responsables para esos tratamientos.

Asimismo, en las distintas etapas pueden intervenir distintos responsables y encargados, además de plantearse situaciones de comunicaciones[51] de datos entre responsables:

Etapa	**Responsable**	**Encargado**
Desarrollo/ Entrenamiento	La entidad que defina los fines del componente IA y decida qué datos se van a emplear para entrenar el sistema. En caso de que se contrate el desarrollo a un tercero, pero este tercero tome las decisiones sobre los datos personales utilizados para entrenar al componente IA para sus propios fines, será considerado responsable la entidad contratada. En el caso que, aquel que defina los fines, adquiera un conjunto de datos personales, será responsable de tratamiento.	La entidad contratada, para entrenamiento o desarrollo, siempre y cuando el contratante fije los términos que definen los fines del tratamiento y las características sustanciales de los datos, tanto si el contratante es quien cede dichos datos como si los obtiene por sí mismo el contratado, y el encargado los utilice sólo para cumplir con los fines del responsable.
Validación	Igual que en el caso anterior.	Igual que en el caso anterior.
Despliegue	En el caso de que la solución IA es un componente[52] que se vende a otra entidad (podría ser formando parte de tratamiento), y ese componente incluye datos de carácter personal, ambas entidades realizan una	La entidad que pone un modelo al servicio de un responsable para que lo explote en un marco de prestación de servicios sin intervenir en esa explotación o que, en caso de hacerlo

[50] Remitimos a la definición de "corresponsable" de este documento y al artículo 26 del RGPD para el caso de que más de una persona tome dichas decisiones.

[51] El artículo 4.2 define como tratamiento también la "…comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión…"

[52] O está incluido en un producto más complejo en el componente de IA es un elemento más.

	comunicación de datos personales y ambas son responsables. Si la comercialización tiene como objeto la venta de un producto que incluya un componente de IA a una persona física para su uso particular, aunque el modelo incluya datos de carácter personal, aplicará la excepción doméstica, salvo que realice un tratamiento para sus propios fines de los datos personales incluidos, en cuyo caso también será considerado responsable.	porque sea necesario para la adecuada ejecución de ese servicio, no utiliza los datos personales para fines propios.
Inferencia/perfilado	La entidad que decide tratar los datos de los interesados con el sistema IA para sus propios fines. Si el tratamiento lo realiza una persona física sobre sus propios datos personales o de aquellas personas en su entorno para una actividad exclusivamente personal o doméstica, se aplicará la excepción doméstica. Esta excepción no aplica a aquellos que proporcionen los medios para tratar datos personales relacionados con tales actividades personales o domésticas[53] para sus propios fines.	Igual que en el caso anterior.
Decisión	La entidad que tome decisiones automatizadas sobre los interesados para sus propios fines.	Igual que en el caso anterior.
Evolución	La entidad que decide tratar los datos de los interesados con el sistema IA, si comunica a una tercera entidad los datos de los usuarios será responsable de la comunicación de datos si no existe una relación de responsable-encargado. La entidad que determina la evolución del componente IA en	En el caso que la entidad que decide tratar los datos de los interesados contrate el tratamiento IA a un tercero, dicho tercero actuará como encargado de tratamiento, siempre que no los trate para sus propios fines.

[53] Considerando 18. El presente Reglamento no se aplica al tratamiento de datos de carácter personal por una persona física en el curso de una actividad exclusivamente personal o doméstica y, por tanto, sin conexión alguna con una actividad profesional o comercial. Entre las actividades personales o domésticas cabe incluir la correspondencia y la llevanza de un repertorio de direcciones, o la actividad en las redes sociales y la actividad en línea realizada en el contexto de las citadas actividades. No obstante, el presente Reglamento se aplica a los responsables o encargados del tratamiento que proporcionen los medios para tratar datos personales relacionados con tales actividades personales o domésticas

	base a los datos de los usuarios, tanto si los datos son cedidos directamente por los interesados como por la entidad que les proporciona servicio, es responsable de dicho tratamiento de evolución o reentrenamiento.	

El modelo de responsabilidad-encargado puede complicarse en el caso de redes cooperativas, tipo "blockchain", que pudieran incorporar modelos de IA. La tabla anterior pretender cubrir los casos más comunes y servir de guía a las nuevas situaciones que puedan surgir en el mercado. Tampoco se entra a detallar las relaciones entre los intervinientes que pueden recoger datos en el marco del Big Data para poner dichos datos en manos de desarrolladores. En particular, no se analizan específicamente modelos de corresponsabilidad.

Hay que ser conscientes que se está tratando el tema de responsabilidad desde el punto de vista de protección de datos, sin ir más allá de otras consideraciones legales o éticas que se deriven de la utilización de la solución de IA.

La decisión de adoptar, en el marco de un tratamiento, una solución técnica basada en IA o en cualquier otra tecnología, es tomada por el responsable, que es quien *"determina los medios y fines del tratamiento"* y es, por tanto, tiene a su cargo la toma de decisión de seleccionar una solución tecnológica u otra. En dicho responsable descansa la obligación de ser diligente a la hora de seleccionar la más adecuada, en particular cuando contrata su desarrollo o la adquiere[54]; exigir y analizar las especificaciones de calidad de la solución; y determinar la extensión del tratamiento y la carga de hacer frente a las consecuencias de sus decisiones. El que toma la decisión de realizar el tratamiento es responsable, y no puede escudarse en la carencia de información o el desconocimiento técnico para evadir su responsabilidad a la hora de auditar y decidir la adecuación del sistema.

Lo que en ningún caso resulta aceptable es trasladar la responsabilidad al propio sistema IA.

[54] Artículo 28.1 del RGPD "Cuando se vaya a realizar un tratamiento por cuenta de un responsable del tratamiento, este elegirá únicamente un encargado que ofrezca garantías suficientes para aplicar medidas técnicas y organizativas apropiados, de manera que el tratamiento sea conforme con los requisitos del presente Reglamento y garantice la protección de los derechos del interesado".

III. CUMPLIMIENTO

El RGPD puede proporcionar una extraordinaria flexibilidad para poder garantizar y demostrar la adecuación de un tratamiento a la norma. Sin embargo, hay un conjunto mínimo de condiciones que deben cumplirse para garantizar la conformidad del tratamiento realizado. Entre ellas pueden citarse:

- La existencia de una base para legitimación del tratamiento de datos personales, (artículos 6 al 11 del RGPD).
- La obligación de informar a los sujetos de los datos y ser transparente (artículos 12 al 14 del RGPD).
- La obligación de proporcionar a los sujetos de los datos mecanismos para el ejercicio de sus derechos (artículos 15 al 23 del RGPD).
- La aplicación del principio de responsabilidad proactiva (artículos 24 al 43) que establecen la necesidad de incorporar una serie de garantías adicionales, más allá de un mínimo, documentadas y orientadas a gestionar el riesgo para los derechos y libertades de los individuos. En particular, la obligación de mantener un registro de actividades de tratamiento (artículo 30 del RGPD).
- El cumplimiento de las condiciones para poder realizar transferencias internacionales de datos (artículos 44 al 50 del RGPD).

Los artículos del RGPD referenciados se desarrollan en la guía Listado para el Cumplimiento Normativo publicada por la AEPD. En este capítulo vamos a comenzar con aquellos aspectos clave que han de tenerse en cuenta a la hora de definir un tratamiento que haga uso de soluciones de IA para garantizar que respeta los principios establecidos en el RGPD. En capítulos posteriores, se repasarán otros aspectos como la responsabilidad proactiva y las transferencias internacionales.

A. LEGITIMACIÓN Y LIMITACIÓN DEL TRATAMIENTO

El establecimiento de una base jurídica legitimadora es el primer paso para determinar el cumplimiento de la solución de IA con el RGPD. La legitimación para las distintas etapas del ciclo de vida y para cada tratamiento se tiene que establecer en la fase de concepción del tratamiento, sea este tratamiento la propia creación de un componente IA o un tratamiento que plantee la utilización de un componente IA. Desde el punto de vista de la Protección de Datos, la legitimación es el primer elemento que hay que establecer dentro de la fase de concepción del tratamiento. Si no se encuentra una base legitimadora no se debe realizar el tratamiento.

En el apartado 1.E de este documento, se han enumerado las distintas etapas del ciclo de vida de una solución IA en las que se podría realizar un tratamiento de datos personales. Cada una de esas etapas tiene un propósito distinto y, además de que puedan tratarse datos del sujeto al que se está prestando el servicio, también pueden llegar a tratarse datos de terceros.

Debido a la naturaleza de los sistemas de IA, en cada etapa del ciclo de vida se podría hacer uso de una base jurídica distinta para:

- El entrenamiento y/o validación del modelo.
- El uso de datos de terceros en la inferencia.
- La comunicación de datos implícitos en el modelo.
- El tratamiento de los datos del interesado en el marco del servicio prestado por la IA.
- El tratamiento de datos del interesado para la evolución del modelo.

El artículo 6 del RGPD establece las seis bases jurídicas por las cuales el tratamiento de datos personales se puede considerar lícito. Las bases jurídicas más habituales que legitimarán el tratamiento en una solución de IA son:

- El tratamiento es necesario para la ejecución de un contrato en el que el interesado es parte, o para la aplicación de medidas precontractuales a petición de este. Podría ser el caso de desarrolladores que contraten a sujetos para hacer uso de sus datos personales en la etapa de entrenamiento del sistema. También podría ser que el responsable del tratamiento, y que proporciona un servicio a terceros interesados que incluye la solución de IA, utilizara los datos de estos en el marco del contrato del servicio.
- El interés legítimo, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño.
- El consentimiento de los interesados, que, como establece el artículo 4.11 del RGPD, es toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen.

Y en ciertos casos más especiales desde el punto de vista de soluciones IA, también pueden ser bases jurídicas:

- Protección de intereses vitales[55].
- Razones de interés público o ejercicio de poderes públicos[56].
- Cumplimiento de obligaciones legales.

Es muy importante tener en cuenta que las dos últimas bases jurídicas han de establecerse vía derecho de la UE o de los Estados miembros, que establecerá la base jurídica del tratamiento. Es decir, un responsable no podrá arrogarse razones, por ejemplo, de interés público si no está establecido en una norma del rango apropiado.

Otro aspecto importante es que se debe tener en cuenta el principio de limitación del tratamiento. Una base jurídica no habilita para el uso de los datos para cualquier propósito y en todo momento[57], sino que debe restringirse a aquellos fines determinados, explícitos y legítimos que se hayan identificado, evitando tratarlos de manera incompatible con esos fines. Además, los interesados cuyos datos son tratados, deben ser conscientes de cómo se van a utilizar, lo que está íntimamente relacionado con el principio de información y transparencia.

La extinción de una base jurídica de legitimación, como puede ser la retirada del consentimiento, no tiene un efecto retroactivo con relación a los resultados obtenidos en un tratamiento ya realizado. Por ejemplo, cuando los datos personales se han empleado para entrenar un componente de IA, la extinción de la base jurídica no invalida la explotación del modelo, aunque el responsable del tratamiento ha de prestar atención a las solicitudes del ejercicio de derechos en materia de protección de datos.

Si el responsable del tratamiento utiliza para el entrenamiento de un componente IA conjuntos de datos de terceros, deberá mostrar la debida diligencia en la comprobación de la legitimidad de la fuente de datos adquirida, incluyendo en el contrato de compra o de prestación del servicio las cláusulas contractuales que reclamen evidencias y compromisos de dicha legitimidad.

---

[55] Artículo 9.2.c del RGPD "el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física, en el supuesto de que el interesado no esté capacitado, física o jurídicamente, para dar su consentimiento".

[56] Como es el caso de Smartcities o control de fronteras.

[57] Con las excepciones señaladas en el artículo 5.1b "…de acuerdo con el artículo 89, apartado 1, el tratamiento ulterior de los datos personales con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos no se considerará incompatible con los fines iniciales" y el artículo 6.4 con relación al tratamiento con fines compatibles por parte del responsable.

Interés legítimo

El Dictamen 06/2014 sobre el concepto de interés legítimo, del Grupo del Artículo 29, desarrolló en detalle cómo evaluar los factores que legitiman el interés del responsable para realizar un tratamiento de datos personales y equilibrarlo con los derechos y los intereses, también legítimos, de los interesados. El interés legítimo es una alternativa de legitimación para tratamientos que requieren, como en algunos casos de ML, acceso a datos de entrenamiento, siempre que se den las circunstancias que permiten su utilización.

Hay que tener en cuenta que el utilizar como base jurídica el interés legítimo reclama del responsable un mayor grado compromiso, formalidad y competencia. Exige realizar una cuidadosa evaluación de que sus intereses legítimos prevalecen sobre el posible impacto en los derechos, libertades e intereses de los interesados. Esta evaluación debe tener en cuenta, entre otras, eventuales medidas compensatorias derivadas de mantener el tratamiento bajo supervisión continua; la adopción de un elevado grado de responsabilidad proactiva; la incorporación de medidas de privacidad por defecto y desde el diseño más estrictas, o la aplicación de buenas prácticas como dar la opción de opt-out a los interesados[58]. En todo caso, el responsable ha de ser capaz de demostrar que dicho impacto no es de tal dimensión como para que no permitir llevar a cabo el tratamiento sobre esa base, debiendo quedar documentado todo este proceso de análisis y toma de decisión en cumplimiento del principio de "accountability".

Si el tratamiento se basa en el interés legítimo no es necesario recabar el consentimiento del interesado, pero las obligaciones de información previstas en los artículos 13 y 14 del RGPD permanecen.

Categorías especiales

Para determinar la base jurídica del tratamiento, es importante tener en cuenta que las categorías especiales de datos, establecidas en el artículo 9 del RGPD, tienen requisitos adicionales para su tratamiento. En estos casos, y antes de analizar una base jurídica que legitime el tratamiento según el artículo 6 del RGPD, es necesario levantar la prohibición previa establecida en el citado artículo 9 en base a alguna de las circunstancias en él contempladas, sin perder de vista las limitaciones adicionales establecidas en la LOPDGDD[59] también en su artículo 9. En particular, debe contemplarse que el consentimiento no levanta la prohibición para tratamientos cuya finalidad principal sea identificar la ideología, afiliación sindical, religión, orientación sexual, creencias u origen racial o étnico del interesado.

Además, el Considerando 71[60] del RGPD establece una restricción adicional sobre el tratamiento de las categorías especiales de datos cuando se pretendan utilizar en decisiones automatizadas y para la elaboración de perfiles, fijando la limitación de que estos solo pueden ser empleados bajo condiciones específicas. En particular, el artículo 22.4 establece que las decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en el interesado o le afecte significativamente de modo similar, no se basarán en las categorías especiales de datos personales salvo que medie el consentimiento del interesado o el tratamiento sea necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

[58] Opinion 7/2015 Meeting the challenges of big data. European Data Protection Supervisor. EDPS
[59] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales
[60] Considerando 71: … Las decisiones automatizadas y la elaboración de perfiles sobre la base de categorías particulares de datos personales únicamente deben permitirse en condiciones específicas.

Estas limitaciones aplican tanto a los datos recogidos de los interesados para el desarrollo o explotación del componente de IA, como a las categorías especiales de datos que se infieran en el curso de los tratamientos.

Tratamientos con fines compatibles

Tal y como establece el artículo 6, apartado 4[61], es posible el tratamiento de datos personales con fines distintos para los que se recogieron inicialmente. La base jurídica para ese nuevo fin se podría basar en un nuevo consentimiento del interesado, en una base legal que lo permitiese, pero también en el caso de que los fines sean compatibles. Por supuesto, el responsable ha de cumplir todos los requisitos para la licitud del tratamiento original y ha de evaluar dicha compatibilidad teniendo en cuenta la relación entre el fin para el que se recogieron los datos y el fin para los que se quiere tratar, el contexto en el que se recogieron, la naturaleza de los datos, prestando especial atención entre otros a las categorías especiales de datos[62], las consecuencias para los interesados y las garantías implementadas en el tratamiento para gestionar el riesgo a los derechos y libertades.

En particular, las operaciones de tratamiento ulterior con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos deben considerarse operaciones de tratamiento lícitas compatibles, con las garantías y excepciones establecidas en el artículo 89 del RGPD.

B. INFORMACIÓN

La información que cada responsable ha de proporcionar a los interesados se establece en los artículos 13 y 14 del RGPD, y el contenido concreto se tendrá que adaptar a la etapa del ciclo de vida de la IA en la que se esté realizando el tratamiento. Como apoyo al cumplimiento de esta obligación, la AEPD ha publicado, con carácter general, la Guía para el Cumplimiento del Deber de Informar así como una nota específica sobre El deber de informar y otras medidas de responsabilidad proactiva en apps para dispositivos móviles. El artículo 11 de la LOPDGDD establece la posibilidad al responsable de ofrecer esta información mediante una aproximación por capas o niveles: una primera capa, de carácter general, con información básica del tratamiento y una segunda capa que completa la información de la primera con mayor nivel de detalle y que sea accesible desde ésta de forma fácil e inmediata, incluso por medios electrónicos:

- En la primera capa deberá consignarse:
  - La identidad del responsable del tratamiento o de su representante.
  - La finalidad del tratamiento.
  - La posibilidad de ejercer los derechos 15 al 22 RGPD.
  - Si el tratamiento incluye la elaboración de perfiles o decisiones automatizadas:
    - Hay que informar claramente que se produce esta circunstancia.
    - Informando de su derecho a oponerse a la adopción de decisiones individuales automatizadas de acuerdo con el art. 22 RGPD.
    - Información significativa sobre la lógica aplicada,
    - Importancia y las consecuencias previstas de dicho tratamiento para el interesado
  - Si los datos personales objeto del tratamiento no han sido obtenidos directamente del afectado, la información básica incluirá también:
    - Las categorías de datos objeto de tratamiento.

[61] El considerando 50 está relacionado con el artículo 6.4.
[62] No sólo a las categorías especiales establecidas en el artículo 9, sino también a los datos personales relativos a condenas e infracciones penales, de conformidad con el artículo 10

- Las fuentes de las que procedieran los datos.

- En la segunda capa, el resto de la información establecida en los artículos 13 y 14 del RGPD.

En el caso de que en el componente IA existan datos personales que puedan ser recuperables hay que informar sobre dicha circunstancia a los interesados y, en su caso, disponer de una base jurídica legitimadora para su comunicación o tratamiento posterior.

Información significativa sobre la lógica aplicada

En el caso de que el interesado esté sometido a decisiones automatizadas o en la elaboración de perfiles a los que hace referencia el artículo 22 del RGPD, un aspecto importante que se establece en el artículo 13.2.f del RGPD es que éste ha de "*disponer de información significativa sobre la lógica aplicada*" y "*la importancia y las consecuencias previstas*".

La palabra "*significativa*", que según la RAE denota "*Que da a entender o conocer con precisión algo*" se ha de interpretar como información que, proporcionada al interesado, le hace consciente del tipo de tratamiento que se está llevando a cabo con sus datos y le proporciona certeza y confianza sobre sus resultados[63].

Cumplir con esta obligación ofreciendo una referencia técnica a la implementación del algoritmo puede ser opaco, confuso, e incluso conducir a la fatiga informativa. Debe facilitarse información que permita entender el comportamiento del tratamiento. Aunque dependerá del tipo de componente IA utilizado, un ejemplo de información que podría tener relevancia de cara al interesado, sería:

- El detalle de los datos empleados para la toma de decisión, más allá de la categoría, y en particular información sobre los plazos de uso de los datos (su antigüedad).
- La importancia relativa que cada uno de ellos tiene en la toma de decisión.
- La calidad de los datos de entrenamiento y el tipo de patrones utilizados.
- Los perfilados realizados y sus implicaciones.
- Valores de precisión o error según la métrica adecuada para medir la bondad de la inferencia.
- La existencia o no de supervisión humana cualificada.
- La referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como la certificación o certificaciones realizadas sobre el sistema de IA. En el caso de sistemas adaptativos o evolutivos, la última auditoría realizada.
- En el caso de que el sistema IA contenga información de terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo.

C. GENERALIDADES SOBRE LOS EJERCICIOS DE DERECHOS

Los responsables que hagan uso de soluciones de IA para tratar datos personales, elaborar perfiles o tomar decisiones automatizadas, han de ser conscientes de que los interesados tienen derechos en el ámbito de la protección de datos que deben ser atendidos.

Por lo tanto, durante la fase de concepción del tratamiento, los responsables han de ser conscientes de que tienen que establecer mecanismos y procedimientos adecuados para poder atender las solicitudes que reciban, y que dichos mecanismos deberán estar adecuadamente dimensionados para la escala del tratamiento que están efectuando.

[63] Está relacionado con el concepto de "explicabilidad" del tratamiento mediante IA.

En el caso de que los datos personales se distribuyan entre una red de responsables, por ejemplo, en el caso de que un modelo IA incluya datos personales, bien en el entrenamiento o en la evolución del sistema, es necesario, siguiendo el principio de responsabilidad proactiva o "accountability"[64], incluir un modelo de gobernanza de la información efectivo que permita la trazabilidad de la información para poder identificar al responsable y hacer posible el ejercicio de dichos derechos. Este modelo de gobernanza de la información debe preverse también cuando en un el tratamiento intervenga un encargado e incluir en el contrato de encargo aquellas tareas que se delegarán a éste con relación a la tramitación de los derechos.

Finalmente, hay que tener en cuenta que en el artículo 11 del RGPD se establece que, a la hora del ejercicio de los derechos de acceso, supresión o limitación del tratamiento, cuando no sea posible la identificación del interesado, el responsable no estará obligado a mantener, obtener o tratar información adicional para identificar al interesado con la única finalidad de cumplir con el RGPD, aunque el afectado tendrá derecho a facilitar información que posibilite su identificación para permitir el ejercicio de los derechos.

D. DERECHO DE ACCESO

El derecho de acceso ha de ejecutarse por el responsable de cada una de las etapas del ciclo de vida de la solución de IA que involucren datos de carácter personal. Esto incluye los datos de entrenamiento que pudieran estar incluidos en los componentes de IA y que puedan ser recuperados por el responsable que explota la solución IA.

E. DERECHOS DE SUPRESIÓN

El derecho de supresión implica una proactividad del responsable del tratamiento para, como establece el Considerando 39[65], garantizar que los datos se suprimen cuando ya no sean necesarios para la finalidad del tratamiento y, en particular, para que se incluyan procedimientos para la revisión periódica del conjunto de datos y plazos para su supresión.

El artículo 17.1 establece la obligación a suprimir sin dilación indebida los datos personales cuando concurren las siguientes situaciones:

- los datos personales ya no sean necesarios en relación con los fines para los que fueron recogidos o tratados de otro modo;
- el interesado retire el consentimiento y el tratamiento no se base en otro fundamento jurídico;
- el interesado se oponga al tratamiento y no prevalezcan otros motivos legítimos
- el interesado se oponga al tratamiento para mercadotecnia directa
- los datos personales hayan sido tratados ilícitamente;
- los datos personales deban suprimirse para el cumplimiento de una obligación legal;
- los datos personales se hayan obtenido en relación con la oferta de servicios de la sociedad de la información

Los datos recogidos para la etapa de entrenamiento, atendiendo a lo señalado en relación con el artículo 11 del RGPD y en cumplimiento del principio de minimización de datos, han de ser depurados de toda la información no estrictamente necesaria para el entrenamiento del modelo.

[64] En particular de las obligaciones establecidas en los artículos 24, 25, 26 y el considerando 66.

[65] Considerando 39: … garantizar que se limite a un mínimo estricto su plazo de conservación. Los datos personales solo deben tratarse si la finalidad del tratamiento no pudiera lograrse razonablemente por otros medios. …, el responsable del tratamiento ha de establecer plazos para su supresión o revisión periódica…

Cuando la etapa de entrenamiento del sistema de IA se ha completado, la organización ha de ejecutar su supresión, a menos que se justifique la necesidad de mantenerlos para el refinado o evaluación del sistema, o se justifique la necesidad y legitimidad de mantenerlos para otras finalidades que resulten compatibles con las que originaron su recogida de acuerdo con las condiciones del artículo 6.4[66] del RGPD y aplicando los principios de minimización de datos. En el caso de que se reciban solicitudes de supresión de los interesados, el responsable tendría que adoptar una aproximación caso por caso, teniendo en cuenta las posibles limitaciones a este derecho previstas en el Reglamento.

En el momento de la distribución de la solución IA, si esta incorpora datos de interesados, será necesario:

- Suprimirlos o, por el contrario, justificar la imposibilidad de hacerlo, en todo o en parte, por la degradación que para el modelo supondría.
- Determinar la base jurídica para llevar a cabo la comunicación de datos a terceros, especialmente si se incluyen categorías especiales de datos.
- Informar sobre dicha circunstancia a los interesados (como se ha señalado anteriormente).
- Demostrar que se han ejecutado las medidas de privacidad por defecto y desde el diseño (sobre todo la minimización de datos).
- En función de los riesgos que podría suponer para los interesados, y teniendo en cuenta el volumen o las categorías de datos, realizar una evaluación del impacto para la protección de datos.

En el caso de que el responsable mantenga los datos del interesado para la personalización del servicio que está ofreciendo la solución de IA, una vez haya extinguido la relación de servicio, estos datos deberán de ser suprimidos.

Limitaciones a la supresión.

El artículo 17.3 del RGPD establece ciertas limitaciones a la supresión. Además, el artículo 32 de la LOPDGDD establece la obligación del responsable a bloquear los datos cuando proceda a su rectificación o supresión.

F. BLOQUEO DE LOS DATOS

El bloqueo es una obligación del responsable con el único propósito de dar respuesta a las posibles responsabilidades derivadas del tratamiento, con el objeto de obtener evidencias de posibles incumplimientos y sólo por el plazo de prescripción de las mismas.

El artículo 32 de la LOPDGDD establece el bloqueo de los datos como la conservación de estos fuera del ámbito del tratamiento, aplicando medidas técnicas y organizativas, que impidan cualquier tipo de proceso, incluyendo su visualización, excepto para la puesta a disposición de los datos a los jueces y tribunales, el Ministerio Fiscal o las Administraciones Públicas competentes, en particular de las autoridades de protección de datos.

Por lo tanto, se ha de tener en cuenta en el diseño de tratamiento como uno de los requisitos. En particular cuando se seleccione o desarrolle una solución IA, será necesario incluir las medidas antes indicadas para bloquear los datos relativos al proceso de inferencia (al menos entradas y resultados obtenidos) que pudieran hacer falta para atender un recurso

[66] Artículo 6.4 Cuando el tratamiento para otro fin distinto de aquel para el que se recogieron los datos personales no esté basado en el consentimiento del interesado o en el Derecho de la Unión o de los Estados miembros que constituya una medida necesaria y proporcional en una sociedad democrática para salvaguardar los objetivos indicados en el artículo 23, apartado 1, el responsable del tratamiento, con objeto de determinar si el tratamiento con otro fin es compatible con el fin para el cual se recogieron inicialmente los datos personales, tendrá en cuenta, entre otras cosas:…

o reclamación de los interesados. Estos mecanismos están también relacionados con los ficheros de registro o logs[67] que se tratarán más adelante.

G. DERECHO DE RECTIFICACIÓN

El responsable tiene la obligación de atender el derecho de rectificación de los datos de los interesados, especialmente aquellos generados por las inferencias y perfiles elaborados por la solución IA.

Por otro lado, en la medida que existan datos inexactos de entrenamiento en el modelo que no tengan un efecto sobre el interesado, por ejemplo, que no haya una posible vinculación de la información inexacta con algún interesado a la hora de distribuir la solución de IA, la inexactitud en los datos puede ser aconsejable como parte de las estrategias de abstracción y ocultación encaminadas a garantizar la aplicación del principio de minimización[68]. Si dichas estrategias conducen a evitar la reidentificación de individuo, haciendo referencia al artículo 11 del RGPD antes citado, no cabría la ejecución del derecho de rectificación.

Por el contrario, si el modelo en sí contiene datos personales inexactos de terceros que pueden ser reidentificados, asociando a dicho terceros una información errónea, es necesario dar respuesta al derecho de rectificación.

H. PORTABILIDAD

El artículo 20 del RGPD establece que, cuando el tratamiento se efectúe por medios automatizados, el interesado tendrá derecho a recibir los datos personales que haya facilitado a un responsable en un formato estructurado, de uso común y lectura mecánica, y a transmitirlos a otro responsable del tratamiento, cuando la legitimación se basa en el consentimiento o el tratamiento es necesario para la ejecución de un contrato.

La existencia de este derecho, desarrollado en las Directrices sobre el derecho a la portabilidad de los datos del Grupo del Artículo 29, clasifica los datos personales en: facilitados, de forma activa y consciente por el interesado; observados del interesado, en virtud del uso del servicio o dispositivo; e inferidos, deducidos o creados por el responsable del tratamiento sobre la base de los datos anteriores. El derecho a la portabilidad aplica a los dos primeros tipos de datos personales.

El responsable de un tratamiento que incluya un componente de IA ha de evaluar y documentar si su tratamiento está obligado a proporcionar la portabilidad sobre los datos facilitados u observados del interesado, en función de lo establecido en el artículo 20 antes citado. En ese caso, el requisito de portabilidad ha de ser tenido en cuenta desde las más tempranas fases de concepción y diseño del tratamiento, en la selección del componente IA y/o por los desarrolladores de componentes IA.

El artículo 20.2 del RGPD establece el derecho a la transmisión de los datos directamente de responsable a responsable, pero sólo cuando sea técnicamente posible. En caso de que existan limitaciones a la portabilidad, es un ejercicio de transparencia informar con antelación a los usuarios de dichas limitaciones.

[67] Ficheros de registro de los eventos que se ejecutan en un sistema.
[68] Como es la aplicación de técnicas de privacidad diferencial

I. TOMA DE DECISIONES BASADAS ÚNICAMENTE EN UN TRATAMIENTO AUTOMATIZADO

Las aplicaciones que ofrecen o soportan un servicio basado en soluciones IA pueden tomar decisiones que afectan a los individuos, sus vidas privadas, su seguridad física, su posición social y su interacción con otras personas.

El RGPD garantiza el derecho a no ser sometido a decisiones automatizadas incluidas la elaboración de perfiles[69] cuando:

- No hay intervención humana. Para que pueda considerarse que existe participación humana, la supervisión de la decisión ha de ser realizada por una persona competente y autorizada para modificar la decisión, y para ello ha de realizar una acción significativa y no simbólica.
- Produzcan efectos jurídicos.
- O afecte de forma similar y significativa al interesado[70].

Se permiten las siguientes excepciones cuando el tratamiento:

- Esté basado en el consentimiento explícito y se apliquen garantías para proteger derechos y libertades.
- Sea necesario para la celebración o ejecución de un contrato, no afecte a categorías especiales de datos y, además, se apliquen garantías para proteger derechos y libertades.
- Esté basado en el derecho de la UE o de España y no afecte a categorías especiales de datos.
- Esté basado en el derecho de la UE o de España y sea necesario para proteger un interés público esencial.

En el apartado “B. Información” antes tratado y, en particular, en el subapartado “Información significativa sobre la lógica aplicada”, así como en el apartado “IV.C Transparencia” se trata sobre los requisitos de información de estos tratamientos.

Cuando la base jurídica sea el consentimiento explícito, el responsable debe diseñar el tratamiento de forma que proteja la libertad de elección de los usuarios. Por un lado, proporcionándole en el momento de solicitarle el consentimiento alternativas viables y equivalentes a la decisión automatizada. Por otro lado, garantizando que si elige no ser objeto a la decisión automatizada no se vaya a introducir un sesgo en la decisión que sea perjudicial para los intereses del afectado. Si no se cumplen estas condiciones, el consentimiento no puede considerarse libre[71]. Estas vías de actuación deben contemplarse desde la misma fase de diseño del tratamiento.

Como buena práctica, y más allá de exigencias derivadas de la protección de datos, la supervisión humana ha de ser una opción en tratamientos basados en IA, y en general, de decisiones automatizadas. Hay que evitar el diseño de sistemas con la orientación “palanca del hombre muerto”[72] y dar siempre la opción de que un operador humano pueda ignorar el algoritmo en un momento dado, procedimentando aquellas situaciones en las que debe optarse por este modo de actuar. Para ello es recomendable documentar las incidencias o

[69] Las decisiones automatizadas pueden llevarse a cabo con o sin elaboración de perfiles; la elaboración de perfiles puede darse sin realizar decisiones automatizadas. No obstante, ambas no son necesariamente actividades independientes. Algo que empieza como un simple proceso de decisiones automatizadas puede convertirse en un proceso basado en la elaboración de perfiles, dependiendo del uso que se dé a los datos. (wp 251)

[70] Dependiendo del caso, podría llegar a considerase que afectan significativamente: el seguimiento de las personas en diferentes sitios web, dispositivos y servicios, alterar las expectativas y deseos de las personas afectadas; alterar la forma en que se presenta un anuncio; o el uso de conocimientos sobre las vulnerabilidades de los interesados. (wp 251)

[71] En “Automated Society Report 2019” se describe un método automático de selección de candidatos basado en el consentimiento que solicita al interesado el acceso a su cuenta de correo electrónico para que un algoritmo evalúe su tráfico de mensajes para obtener un perfil como futuro empleado. Si no se plantea alternativa, o la mera presentación de un CV penaliza el proceso de selección, el consentimiento no sería válido.

[72] Dead man switch

los cuestionamientos de las decisiones automáticas recibidas de los interesados, de modo que, de su análisis, sea posible detectar situaciones en las que es necesaria la intervención humana porque el tratamiento puede no estar funcionando de la manera esperada.

Con relación al tratamiento de datos de menores, el Considerando 71 establece que no se les deben aplicar las decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles[73], con efectos jurídicos o significativamente similares. Sin embargo, al no estar reflejada esta prohibición en el articulado, no se considera que la prohibición tenga un carácter absoluto, pudiendo darse excepciones cuando resulte imprescindible para proteger el bienestar del menor y se implementen las garantías adecuadas.

[73] El Dictamen 02/2013 sobre las aplicaciones de los dispositivos inteligentes (WP 202) del GT29, adoptado el 27 de febrero de 2013, en la sección 3.10 específica sobre los niños, especifica en la página 32 que «los responsables del tratamiento no deben procesar los datos sobre niños con fines de publicidad comportamental, ni directa ni indirectamente, por quedar esto fuera del alcance de comprensión del niño y, por tanto, exceder de los límites de tratamiento lícito»

IV. GESTIÓN DEL RIESGO PARA LOS DERECHOS Y LIBERTADES

La gestión del riesgo para los derechos y libertades es una actividad continua y forma parte del concepto de Responsabilidad Proactiva o "accountability" que se establece en el RGPD.

Cuando se tratan datos personales y sea proporcional con relación a las actividades de tratamiento, la organización ha de sistematizar las medidas de cumplimiento en una Política de Protección de Datos, como se establece en el artículo 24 del RGPD. Esta política convivirá con una política de calidad, una política de sistemas de información, una de seguridad y una política de toma de decisiones, entre otras.

El RGPD no establece de forma específica qué elementos de responsabilidad proactiva hay que implementar de forma obligatoria o la forma de implementarlos, limitándose a proporcionar un amplio grado de flexibilidad para que muy distintos tratamientos se adecúen a la norma. Sin embargo, sí establece que dichas medidas se han de seleccionar siguiendo un análisis basado en el riesgo (RBT[74]), y específicamente en el riesgo para los derechos y libertades de las personas con relación al tratamiento de sus datos personales y la elaboración de perfiles[75].

El RBT tiene dos fases fundamentales: la primera es la identificación de amenazas y evaluación del nivel de riesgo intrínseco existente; y la segunda es la gestión de este riesgo mediante la implementación de medidas técnicas y organizativas, adecuadas y proporcionales, para eliminarlo o al menos mitigarlo, reduciendo el impacto o la probabilidad de que se materialicen las amenazas identificadas. Por último, una vez implementadas las medidas seleccionadas, se debe evaluar el riesgo residual[76] remanente y mantenerlo controlado.

Tanto para determinar el riesgo, como para establecer las medidas apropiadas para gestionarlo, es preciso realizar un análisis del tratamiento, dividiendo el mismo en sus distintas fases y administrar las peculiaridades de cada una de ellas.

Para determinar el nivel de riesgo de un tratamiento basado o que contiene fases en las que existe un componente IA hay que tener en cuenta:

- Los riesgos que se derivan del tratamiento en sí mismo, siendo el más característico el que se deriva del sesgo en los sistemas de toma de decisiones sobre las personas o su discriminación (*algorithmic discrimination*[77]).
- Los riesgos que se derivan del tratamiento con relación al contexto social y los efectos colaterales que se puedan derivar de él, indirectamente relacionados con el objeto de tratamiento.

A. EVALUACIÓN DEL NIVEL DE RIESGO

Evaluar el nivel de riesgo de un tratamiento es una operación necesaria para determinar el modo en que se van a aplicar las medidas tendentes a suprimirlo o minimizarlo, en particular cuando el riesgo detectado es alto según las condiciones establecidas en la norma, fundamentalmente en:

- El artículo 35.3 del RGPD.

[74] Risk Based Thinking

[75] Existen muchos tipos de riesgo: de continuidad de negocio, fraude, financiero, de imagen, de oportunidad, de fiabilidad de la tecnología, de seguridad, etc.

[76] La ISO 31000 define riesgo residual como aquel riesgo que subsiste, después de haber implementado los controles. Controles son todas aquellas medidas preventivas, detectivas o correctivas destinas a minimizar el riesgo.

[77] Supone trasladar a la implementación del componente IA los prejuicios existentes en la sociedad vinculados con categorías espaciales de datos u otros atributos, como nacionalidad. https://eticasfoundation.org/wp-content/uploads/2018/03/IG_AlgoritmicAFinal.pdf.

- El artículo 28.2 de la LOPDGDD.
- Las Listas de tipos de tratamientos de datos que requieren EIPD (art 35.4) publicadas por la AEPD.

En los artículos y listas anteriores se establecen determinadas condiciones que se han de cumplir en un tratamiento, en ocasiones de forma simultánea, para que sea considerado de alto riesgo. Las listas anteriores no han de verse como exhaustivas, sino como una guía para el análisis del riesgo. En cualquier caso, el responsable ha de ser capaz de identificar los riesgos adicionales derivados de la novedad de la solución IA y su modo de implementación como, por ejemplo, en el caso de que el componente IA que se distribuya incluya datos personales. Hay que recalcar que este análisis trasciende los límites de la entidad responsable, y está orientado a medir el impacto que la actividad realizada pueda tener en el entorno social en el que se realiza o despliega la aplicación.

El responsable del desarrollo, mantenimiento y/o distribución de un componente IA, así como el responsable de un tratamiento que incluya componentes IA, ha de tomar, en cada una de las respectivas etapas y responsabilidades, las medidas oportunas para minimizar o eliminar los factores de riesgo.

B. LA EVALUACIÓN DE IMPACTO DE LA PRIVACIDAD - EIPD

La EIPD es una obligación establecida en el RGPD cuando los niveles de riesgo asociados al tratamiento son elevados. Esta obligación implica ir más allá de realizar la mera gestión del riesgo del tratamiento, puesto que exige una formalidad adicional a la hora de ejecutar dicha gestión.

La necesidad de que cada responsable lleve a cabo una evaluación de impacto de la protección de datos se establece en el artículo 35 del RGPD cuando, según el apartado 1, "*el tratamiento entrañe un alto riesgo para los derechos y libertades de las personas físicas*".

En particular, pero no de forma exclusiva, y tal como establece el artículo 35.3.a[78], es necesario realizar una EIPD cuando se realice la elaboración de perfiles, basados en tratamientos automatizados (pero no necesariamente exclusivamente automatizados[79]), sobre los que se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente.

La EIPD se ha de realizar antes de que se ejecute el tratamiento efectivo de los datos de carácter personal, es decir, antes de iniciar el tratamiento[80]. Por lo tanto, la validación ha de realizarse antes del diseño/selección e implementación de la solución IA para un determinado tratamiento y que, de esta forma, sea posible identificar cuáles son los requisitos de privacidad a incorporar y poder aplicar, de manera efectiva, las medidas de privacidad desde el diseño y por defecto.

El artículo 35.2 del RGPD establece la obligación del responsable de recabar el asesoramiento del delegado de protección de datos, si éste ha sido nombrado.

También es importante tener en cuenta que el artículo 35.9[81] del RGPD establece que, cuando proceda, se recabará la opinión de los interesados.

[78] Artículo 35.3.a evaluación sistemática y exhaustiva de aspectos personales de personas físicas que se base en un tratamiento automatizado, como la elaboración de perfiles, y sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente de modo similar.

[79] Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679.

[80] Artículo 35.1.… el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales.

[81] Artículo 35.9 Cuando proceda, el responsable recabará la opinión de los interesados o de sus representantes en relación con el tratamiento previsto, sin perjuicio de la protección de intereses públicos o comerciales o de la seguridad de las operaciones de tratamiento.

En el caso de las soluciones IA, esta opinión es particularmente importante para entender, además de la tecnología en sí, el contexto preciso en el que se va a emplear esta y los modelos específicos de riesgo para los derechos y libertades de los interesados. El concepto de interesado ha de entenderse extendido tanto a los operadores humanos que interpretan o supervisan los resultados de la IA como a los sujetos sometidos a su tratamiento.

En particular, cuando el tratamiento, de forma automatizada, elabore perfiles y tome decisiones, hay que identificar todas estas decisiones en las distintas fases del tratamiento, detallarlas, analizar los parámetros de funcionamiento, como los márgenes de error, y evaluar cuidadosamente qué efectos tienen sobre los interesados.

La EIPD debe estar documentada y, cuando se evidencie un alto riesgo residual, ha de someterse a consulta de la autoridad de control con las condiciones establecidas en el artículo 36 del RGPD.

Una EIPD ha de concretarse en la adopción de una serie de medidas específicas y concretas para la gestión del riesgo, algunas de ellas orientadas a reforzar las obligaciones de cumplimiento en función de dicho riesgo y que afectan a:

- La concepción del tratamiento, en sus fases, procedimientos, tecnologías y extensión.
- La incorporación de medidas de privacidad por defecto y desde el diseño en el tratamiento y que sigan los principios de:
    - **Minimizar** la cantidad de datos que son tratados, tanto en volumen de información recopilada como en el tamaño de la población de estudio, así como a lo largo de las diferentes fases del tratamiento.
    - **Agregar** los datos personales en la medida de lo posible para reducir al máximo el nivel de detalle que es posible obtener.
    - **Ocultar** los datos personales y sus interrelaciones para limitar su exposición y que no sean visibles por partes no interesadas.
    - **Separar** los contextos de tratamiento para dificultar la correlación de fuentes de información independientes, así como la posibilidad de inferir información.
    - Mejorar la **Información** a los interesados, en tiempo y forma, de las características y bases jurídicas de su tratamiento para fomentar la transparencia y permitir a los interesados tomar decisiones informadas sobre el tratamiento de sus datos.
    - Proporcionar medios a interesados para que puedan **controlar** cómo sus datos son recogidos, tratados, usados y comunicados a terceras partes mediante la implementación de mecanismos adaptados al nivel de riesgo que les permita realizar el ejercicio de sus derechos en materia de protección de datos.
    - **Cumplir** con una política de privacidad compatible con las obligaciones y requisitos legales impuestos por la normativa.
    - **Demostrar**, en aplicación del principio de responsabilidad proactiva, el cumplimiento de la política de protección de datos que esté aplicando, así como del resto de requisitos y obligaciones legales impuestos por el Reglamento, tanto a los interesados como a las autoridades de control. Esto implica auditar dinámicamente el resultado/ las conclusiones de los tratamientos, evaluando las divergencias o desviaciones sobre los inicialmente previstos o evaluados como previsibles, incluidos los algoritmos ejecutados, para adoptar, en su caso, medidas correctivas, incluida, la supresión de la información y documentar detalladamente el análisis realizado y las medidas adoptadas.

- La identificación de requisitos de seguridad que minimicen el riesgo para la privacidad.
- La adopción de medidas específicas dirigidas a implementar un sistema de gobernanza de los datos personales que permitan demostrar el cumplimiento de principios, derechos y garantías para gestionar el riesgo de los tratamientos realizados.

La AEPD proporciona guías para abordar los aspectos anteriores de forma concreta para cualquier tipo de tratamiento genérico, destacando la Guía práctica para las evaluaciones de impacto en la protección de datos personales, la Guía de Privacidad desde el Diseño, Modelo de informe de Evaluación de Impacto en la Protección de Datos para Administraciones Públicas, así como herramientas que ayudan a realizar la EIPD, como GESTIONA.

Antes de repetir lo que se puede encontrar en dichas guías, a las que nos remitimos, se van a desarrollar algunos aspectos que pueden ser más determinantes para los tratamientos que empleen soluciones de IA.

C. TRANSPARENCIA

Según el Considerando 78[82], el principio de transparencia es una medida de privacidad por defecto para permitir, entre otros, que los interesados puedan supervisar el tratamiento al que están sometidos.

El principio de transparencia se desarrolla en los Considerandos 39 y 58. En estos Considerandos se interpreta la obligación de información a los interesados de un modo que va más allá de lo dispuesto en la letra de los artículos 13 y 14 del RGPD. En particular, los Considerandos comentan la obligación de que "*toda información y comunicación relativa al tratamiento de dichos datos sea fácilmente accesible y fácil de entender*", "*sea concisa*", "*se utilice un lenguaje sencillo y claro*", que "*en su caso, se visualice*", que "*podría facilitarse en forma electrónica*", que se proporcione "*información añadida para garantizar un tratamiento leal y transparente*" y que los interesados "*deben tener conocimiento de los riesgos, las normas, las salvaguardias*" del tratamiento.

En el caso de tratamientos basados en IA, la transparencia puede ser considerada un aspecto crítico. Debe permitir a los interesados ser conscientes del impacto que el empleo de dichas soluciones puede llevar asociado. De ahí que la transparencia esté dirigida tanto a los interesados como a los operadores del tratamiento. En particular, la transparencia está ligada con una información veraz sobre la eficiencia, las capacidades y las limitaciones reales de los sistemas de IA, que evite la creación de falsas expectativas, en los usuarios y los interesados, que puedan ocasionar una mala interpretación de las inferencias que se realizan en el marco del tratamiento. También, la transparencia está ligada a información sobre el contexto y situación del tratamiento, como la existencia de terceras partes, la localización física/virtual de la solución AI, etc.

La transparencia no se reduce a un instante puntual, sino que debe ser entendida como un principio en torno al que orbita de forma dinámica el tratamiento realizado y que afecta a todos y cada uno de los elementos y participantes que intervienen en la solución.

Durante la etapa de entrenamiento

Durante la etapa de entrenamiento, si se utilizan datos de carácter personal, hay que informar claramente al interesado sobre la posibilidad de que sea reidentificable a partir de

[82] Considerando 79 …, el responsable del tratamiento debe adoptar políticas internas y aplicar medidas que cumplan en particular los principios de protección de datos desde el diseño y por defecto. Dichas medidas podrían consistir, entre otras, …dar transparencia a las funciones y el tratamiento de datos personales, permitiendo a los interesados supervisar el tratamiento de datos y al responsable del tratamiento crear y mejorar elementos de seguridad.

los datos del modelo o, como se establece en el artículo 11.2 del RGPD, informarle de que no es posible realizar dicha reidentificación.

Certificación

Uno de los problemas que se plantean a la hora de proporcionar transparencia con relación a los algoritmos de IA es la cuestión de la confidencialidad necesaria para preservar la propiedad industrial.

Este problema no es exclusivo de la IA, sino que es común a cualquier desarrollo tecnológico, como, por ejemplo, los módulos criptográficos en sistemas seguros, o los circuitos integrados de las tarjetas de crédito. Sin embargo, en los sectores en los que se emplean estas soluciones es imperativo no confiar únicamente en las manifestaciones de fabricantes y distribuidores, sino que obligatoriamente se establecen mecanismos basados en terceros confiables para determinar los niveles necesarios de calidad y confiabilidad. Estos mecanismos se materializan bien dando acceso a las autoridades de control a los aspectos internos de dichas soluciones, o bien estableciendo mecanismos de certificación por terceros independientes[83].

De igual forma, cuando la IA evoluciona de ser un elemento experimental a un producto, es necesario incorporar las mismas garantías que se le reclaman a cualquier otro servicio tecnológico. El desarrollo y aplicación de esquemas de certificación que establezcan marcos de referencia (para fabricantes, responsables, autoridades y usuarios) que permitan acreditar por un tercero independiente que tanto la solución IA en particular, así como que la implementación del tratamiento en general, cumplen con lo establecido en el RGPD y lo desarrollado en esta guía, es un elemento capital para la aplicación del principio de transparencia.

El RGPD establece en el artículo 42 la posibilidad de desarrollar mecanismos de certificación específicos en materia de protección de datos y de sellos y marcas de protección de datos como herramientas para demostrar el cumplimiento del RGPD. Estos esquemas pueden cubrir distintos aspectos del empleo de la IA y ser herramientas para demostrar diligencia. Tal y como explica el Considerando 100, los mecanismos de certificación sirven también para aumentar la transparencia en los tratamientos de datos de carácter personal.

Decisiones automatizadas y elaboración de perfiles

Los responsables deben proporcionar a los interesados información suficiente respecto al tratamiento al que están siendo sometidos, así como información sobre los mecanismos que les permita solicitar intervención humana en la evaluación o cuestionar la decisión tomada por el sistema de forma automática, más allá de lo estrictamente establecido en los artículos 13 y 14 del RGPD y que se ha descrito en el apartado “III B. Información” de este documento.

Personal del responsable

Los responsables que adquieren este tipo de soluciones y sistemas deben proporcionar información precisa y formación específica a su personal sobre las limitaciones del sistema de IA.

Cuando el tratamiento sea una herramienta de ayuda a la toma de decisión, es necesario adoptar medidas para gestionar el riesgo de que el elemento humano se comporte como

[83] Como es el caso de la certificación basada en Common Criteria o CC, y donde se pueden encontrar listas de productos certificados https://www.commoncriteriaportal.org/

una mera correa de transmisión de las inferencias realizadas por una solución IA. Estas medidas incluyen información al operador (como se ha indicado anteriormente), formación y auditorías de su comportamiento.

Por otro lado, hay que prevenir errores de interpretabilidad por parte de los operadores. Los valores inferidos han de representarse de forma que reflejen la realidad de la inferencia y sus límites a los operadores humanos o a las fases posteriores del tratamiento. A la hora de la explotación del sistema, es necesario ofrecer información en tiempo real al operador o usuario final de los valores de exactitud y/o calidad de la información inferida en cada momento. Cuando la información inferida no alcance unos umbrales mínimos de calidad, se ha de indicar, de forma explícita, que dicha información es nula o no tiene ningún valor[84].

En el caso de que las soluciones sean adquiridas a un tercero, este ha de proporcionar información suficiente al responsable para que pueda gestionar estos riesgos, así como sobre la mejor forma de hacerlo.

El Delegado de Protección de Datos como herramienta de transparencia

El nombramiento de un DPD es una de las mejores medidas que puede adoptar el responsable para orientar la implementación de las políticas de transparencia, en particular, para gestionar un canal de información a los interesados. El usuario puede obtener así información sobre la solución de IA (bien del desarrollo de la IA, de su mantenimiento o de su explotación en el marco de un tratamiento) directamente del responsable.

El artículo 37.1 del RGPD como el artículo 34 de la LOPDGDD establecen las condiciones que obligan a un responsable/encargado a contar con un Delegado de Protección de Datos (DPD), bien por la naturaleza del tratamiento o por el tipo de actividad.

El utilizar en el tratamiento una solución de IA no implica, *per se*, una obligación de tener un DPD, aunque hay dos casos, detallados en las letras b) y c) del anterior artículo 37.1, que sí determinarán la obligación de contar con su presencia en la organización: la observación habitual y sistemática a gran escala y el tratamiento de categorías especiales de datos o datos relativos a condenas e infracciones penales también a gran escala.

Según las Directrices sobre los delegados de protección de datos publicadas por el WP29[85], el DPD es un elemento clave para garantizar el cumplimiento y la gestión del riesgo para los derechos y libertades de los interesados.

Por tanto, el DPD, aun en los casos que no sea obligatorio, puede ser de gran utilidad en aquellas entidades que emplean soluciones basadas en IA y que tratan datos personales, o que desarrollan soluciones de IA que hacen uso de datos personales para el entrenamiento de los modelos. El DPD se convierte en un elemento clave para poder gestionar el riesgo y aplicar de forma efectiva los mecanismos de responsabilidad proactiva. En particular, el artículo 35 identifica al DPD como un rol fundamental en la realización de la EIPD y una de las herramientas para implementar la transparencia de cara al usuario.

D. EXACTITUD

El término “exactitud” se define en el artículo 5.1.d del RGPD como uno de los principios de protección de datos de carácter personal. En el caso de tratamientos que incorporan

[84] Por ejemplo, en el caso de sistemas que detecten comportamiento corporal sospechoso, no debería señalarse en la pantalla del operador un sujeto simplemente porque no hay ningún otro con una valoración más alta de riesgo.

[85] “El DPD desempeña un papel fundamental en la promoción de una cultura de protección de datos dentro de la organización y contribuye a la aplicación de elementos esenciales del RGPD, como los principios relativos al tratamiento de datos, los derechos de los interesados, la protección de los datos desde el diseño y por defecto, el registro de las actividades de tratamiento, la seguridad del tratamiento y la notificación y comunicación de las violaciones de la seguridad de los datos”

soluciones IA, la existencia de sesgos en los modelos de inferencia está íntimamente ligado con la exactitud o calidad del dato[86].

Según el segundo párrafo del Considerando 71 del RGPD, los datos asociados a los interesados[87], ya sean los datos directamente recogidos o los inferidos, han de ser exactos. En particular, se hace explicito que el responsable del tratamiento ha de utilizar "*procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles*" que garanticen que los datos vinculados con el interesado son exactos. Es obligatorio demostrar y documentar que los procedimientos empleados para la inferencia de información sobre un interesado son precisos y, por tanto, estables y predecibles.

Cuando un responsable selecciona el empleo de un componente de IA para un tratamiento ha de tener la diligencia de asegurarse que los datos tratados, generados y vinculados con el interesado cumplen con las condiciones anteriores.

Factores que influyen en la exactitud

Respecto a que los datos inferidos sobre el interesado sean exactos, hay tres factores que pueden influir sobre esa exactitud:

- La propia implementación del sistema IA. Existen IA's, como por ejemplo sistemas expertos basados en reglas, en los que la concepción del propio sistema puede introducir errores que conduzcan a inferencias erróneas; o sistemas basados en ML que no son capaces de modelar el tratamiento deseado. Se pueden producir errores porque elementos externos a la IA, como por ejemplo lectores biométricos, introduzcan errores en los datos de entrada. Por otro lado, existe la posibilidad de que existan errores de programación o de diseño que trasladen, de forma errónea, el modelo a su implementación práctica. En estos casos se puede decir que el que sesgo está "hardwired" por las decisiones que se toman cuando se construye el modelo de análisis (sesgos de evaluación y agregación).
- El conjunto de datos de entrenamiento o validación está viciado con errores, información deliberadamente errónea[88] o sesgos que imposibilitan que las inferencias sean correctas. Estos sesgos podrían ser inherentes, si ya existen en los datos que alimentan el sistema, como mala calidad de los datos, datos ausentes, o muestreo selectivo. También podrían ser de representación y medida, debidos a cómo se da formato al conjunto de datos para alimentar al sistema.
- La evolución sesgada del modelo IA. En el caso de IA que implementan técnicas adaptativas, la solución IA puede estar siendo utilizada mayoritariamente por un grupo de sujetos con características particulares que introducen nuevos sesgos de realimentación.

En cuanto a la exactitud requerida en caso de datos de entrenamiento, se han de emplear métricas y técnicas de depuración y trazabilidad para garantizar la fidelidad e integridad del conjunto de datos. Aunque no es común a todas las soluciones IA, en algunos modelos los datos de entrenamiento y explotación pueden clasificarse en datos "duros" y datos "blandos". Los primeros son datos objetivos, generalmente cuantificables, como, por ejemplo: calificaciones, porcentaje de asistencia, valores analíticos, resultados de pruebas, etc. Los datos "blandos" son aquellos datos generalmente cualitativos que contienen un componente subjetivo o de incertidumbre. Ejemplos de datos "blandos" podrían ser: los obtenidos a través

[86] El término "calidad del dato" aparece en el RGPD en el artículo 47.2.d

[87] Todos los interesados tienen los mismos derechos respecto a la protección de datos, no sólo los de grupos calificados como "de riesgo", y no cabe una "discriminación positiva".

[88] La información errónea puede aparecer bien por un ataque realizado sobre el conjunto de datos o bien por la inclusión de puertas traseras por la manipulación consciente de los datos.

de procesamiento de lenguaje natural, opiniones, evaluaciones personales, encuestas, etc. Aunque los datos “duros” no están exentos de errores y sesgos, el responsable ha de tener especial cuidado en evaluar los problemas de exactitud que se pueden derivar de utilizar o dar mayor relevancia a los datos “blandos” como fuente de información.

El sesgo no es un problema exclusivo de un sistema de IA, sino que puede aparecer en cualquier sistema de proceso automático/no automático que toma decisiones o que realiza elaboración de perfiles. Existen técnicas conocidas como Algorithmic Impact Assessment[89] (AIA), orientadas a examinar y determinar la posible existencia de sesgos en los algoritmos utilizados en las soluciones de IA y a garantizar la equidad en la implementación del modelo. Estas métricas han de analizar la lógica implementada, para que la misma no produzca inexactitudes por diseño, y emplear modelos maduros de pruebas y test de la lógica para detectar errores de diseño[90].

Información biométrica

La exactitud es particularmente crítica cuando el tratamiento está basado en información biométrica, como IA sobre reconocimiento facial, huellas dactilares, voz, etc. En ese caso, se han de tener en cuenta factores de rendimiento (falsos positivos, falsos negativos y otros) y también el impacto sobre la recogida de los datos de personas con alguna discapacidad o singularidad física. Especialmente, cuando la solución IA es incapaz de identificar al individuo como tal[91], se produce un perfilado erróneo incluso antes de la ejecución del tratamiento principal[92].

El responsable ha de tener en cuenta que, aunque dichos usuarios pueden ser una minoría, se han de establecer mecanismos alternativos para evitar la exclusión de un sujeto porque la solución de IA no es capaz de procesar las características biométricas de los interesados[93].

Combinación de perfilados

Puede darse el caso que el componente IA elabore perfiles o decisiones sobre un mismo individuo, pero en distintas situaciones o para distintos objetivos. Por ejemplo, podrían existir tratamientos en el ámbito penal que utilizan soluciones IA que permiten evaluar un perfil sobre la posibilidad de que un individuo no se presente a un emplazamiento ante el tribunal y, simultáneamente, un perfil sobre el riesgo que el mismo interesado tiene de reincidir en el delito. En ese caso, y a menos que se justifique documentalmente, es conveniente que estos perfiles se elaboren de forma independiente, sin que los valores o el riesgo inferido para uno de ellos influya en la inferencia realizada para el otro.

---

[89] Los Algorithmic Impact Assessment son herramientas para permitir evaluar a los interesados que los componentes de IA cumplen con determinados parámetros de calidad y que son adecuados para la tarea encomendada. Referencias sobre los mismos se pueden encontrar, por ejemplo, en https://ainowinstitute.org/aiareport2018.pdf

[90] Trabajos recientes han planteado preocupaciones sobre el riesgo de sesgos no intencionales en estos modelos, que afectan negativamente a individuos de ciertos grupos. Si bien se han propuesto muchas métricas de sesgo y definiciones de equidad, no hay consenso sobre qué definiciones y métricas deberían usarse en la práctica para evaluar y auditar estos sistemas. No obstante, se han desarrollado diferentes métricas orientadas a evaluar la discriminación algorítmica como Aequitas, un kit de herramientas de auditoría de sesgos de código abierto desarrollado por el Center for Data Science and Public Policy.

[91] En el estudio Discrimination, artificial intelligence, and algorithmic decision-making publicado por el Consejo de Europa, se enumeran los siguientes ejemplos: el software de seguimiento facial de Hewlett Packard no reconoció los rostros oscuros como caras, la aplicación Google Photos etiquetó una foto de una pareja afroamericana como "gorilas", una cámara Nikon seguía preguntando a personas de origen asiático: ¿alguien parpadea?, un hombre asiático tuvo su foto de pasaporte rechazada, automáticamente porque "los ojos del sujeto están cerrados", Buolamwini y Gebru descubrió que "las mujeres de piel más oscura son el grupo más mal clasificado (con error tasas de hasta el 34,7%). La tasa de error máxima para los hombres de piel más clara es del 0,8% ".

[92] Existen casos de personas con huellas dactilares poco marcadas que sometidas a sistemas de acceso mediante identificación por huella se encuentran con continuos problemas de acceso.

[93] En definitiva, para evitar una discriminación por no ser “biométricamente adecuado”

Verificación vs. Validación

La realización de pruebas y/o verificación del componente de IA es una parte capital en el desarrollo de este, así como como un criterio importante para seleccionar uno u otro componente por el responsable del tratamiento. El responsable ha de tener en cuenta que la inclusión en un tratamiento de un componente IA verificado no implica que el conjunto del tratamiento esté validado, ni tampoco la adecuación del componente IA para ese tratamiento específico. Las pruebas del componente IA aseguran que los resultados del diseño y desarrollo cumplen con los requisitos del componente. La validación del tratamiento llega más lejos, ya que se asegura de que los productos y servicios resultantes satisfacen los requisitos relativos a una aplicación específica o uso previsto[94]. La validación garantiza que el tratamiento, así como la solución IA sobre la que se apoya, cumple con los resultados planificados para un producto o servicio concreto.

Es decir, la validación del tratamiento que incluye un componente IA ha de realizarse en las condiciones que reflejen el entorno real en el que se pretende desplegar el tratamiento[95]. A su vez, el proceso de validación exige una revisión periódica, en la medida que evolucionen dichas condiciones o el tratamiento en sí.

Garantía de exactitud como un proceso continuo

En el caso de que las soluciones IA evolucionen, sobre todo cuando se realimenten tanto con su interacción con el interesado como con la interacción con terceros distintos del interesado, es necesario realizar procesos de recalibración del modelo.

La deriva en la exactitud del perfilado realizado por una solución de IA debido al efecto "filtro burbuja"[96] puede realimentar los sesgos que, sobre sí mismo o sobre terceros, tiene el usuario del tratamiento que incluye la IA[97].

E. MINIMIZACIÓN

La minimización se define implícitamente en el artículo 5.1.c del reglamento como el proceso dirigido a garantizar que los datos personales son adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados. Según el Considerando 59, los datos personales solo deben tratarse si la finalidad del tratamiento no pudiera lograrse razonablemente por otros medios.

Los interesados, las categorías de datos, y el periodo de conservación de los datos que se pueden procesar en un tratamiento están ligados a la legitimación de ese tratamiento.

La minimización persigue optimizar el tratamiento desde el punto de vista de la protección de datos, analizando las necesidades de proceso de datos en las distintas fases, y cumpliendo los requisitos anteriores. Consiste en:

[94] ISO 9001:2015 Sistemas de gestión de la calidad. Requisitos.

[95] En el "Manual Práctico de Inteligencia Artificial en Entornos Sanitarios" referenciado en la bibliografía, se describe el caso en el que un sistema para el diagnóstico pediátrico de neumonía obtuvo en su desarrollo una precisión del 97%. Sin embargo, cuando se aplicó dicho algoritmo con una población de Madrid la precisión bajó al 64%. El análisis de los datos de entrenamiento evidenció que la población utilizada tenía una edad de 0 a 5 años, mientras que el ámbito de aplicación de tratamientos pediátricos en Madrid abarca a niños hasta los 14 años

[96] Filtro burbuja o cámara eco es una situación muy común en los buscadores de contenidos, que aprenden de los gustos del usuario y sólo ofrecen el contenido que creen que es del gusto del usuario, obviando cualquier otra opción y provocando que el usuario esté encerrado en una "burbuja" que no le permite acceder a todas las posibles opciones.

[97] Por ejemplo, una IA que evalúe el estado psicológico del sujeto, puede realimentar la visión que tiene el individuo sobre sí mismo, conduciéndolo a una deriva de su percepción.

- Limitar la extensión de las categorías de datos[98] que se utilizan en cada fase del tratamiento a aquellas que son estrictamente necesarias y relevantes.
- Limitar el grado de detalle o precisión de la información[99], la granularidad de la recogida en tiempo y frecuencia y la antigüedad de la información utilizada.
- Limitar la extensión en el número de interesados de los que se tratan los datos.
- Limitar la accesibilidad de las distintas categorías de datos al personal del responsable/encargado o incluso al usuario final (si hay datos de terceros en los modelos de IA) en todas las fases del tratamiento.

Es necesario que en la concepción de la solución IA se estudie la forma de implementar estas restricciones, describiendo y analizando el ciclo de vida de los datos a lo largo de todas las etapas del tratamiento. El análisis no puede centrarse únicamente en la solución IA desde el punto de vista técnico, sino que debe englobar todo el tratamiento en el que se incluye y alcanzando tanto los aspectos automatizados como los no automatizados de todas y cada una de las fases del tratamiento.

Datos de entrenamiento

En el caso de ML, es necesario balancear la necesidad de datos para el entrenamiento de los sistemas con los riesgos para los derechos y libertades de los interesados. El grado de calidad de los datos de entrenamiento no se mide simplemente por la acumulación de datos, sino por los parámetros de relevancia, actualidad, fiabilidad, robustez[100] y extensión de la tipología de datos a los ámbitos relevantes del tratamiento. Para aplicar dicho criterio de proporcionalidad es recomendable consultar a profesionales con conocimientos sobre la ciencia de datos[101], disciplina que vas más allá de la especialización en algoritmos de ML. También habrá que escuchar a expertos con conocimiento sobre los principios de protección de datos ,especialistas en la lógica de negocio y al delegado de protección de datos, cuando esté designado[102].

Técnicas de minimización

Existen diferentes técnicas de minimización de datos[103] para las aplicaciones de Inteligencia Artificial, y algunas de ellos son específicas para ML. Las técnicas están en continuo desarrollo, pero podemos citar:

- Realización de un análisis previo de las condiciones que han de cumplir los datos para que sean considerados de alta calidad y con una gran capacidad predictora para la aplicación concreta.
- Análisis de forma crítica de la extensión de la tipología de datos empleados en cada etapa de la solución IA.
- Supresión de datos de datos no estructurados, o información no necesaria recogida durante el preproceso de la información.

[98] Extensión de la categoría de datos se refiere a el número de campos de datos asociados a una persona física: nombre, direcciones físicas y lógicas, campos sobre su salud, situación laboral, social, relaciones, gustos, creencias, ideología, … Cuantos más campos, con más precisión y mayor diversidad se recojan de un individuo tanto mayor será la extensión de las categorías de datos tratadas.

[99] Algo tan sencillo como la fecha de nacimiento se puede precisar desde identificando la década, hasta precisar el año, día y hora.

[100] En cuanto sea aplicable al entrenamiento para distintos tratamientos.

[101] La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático, y la analítica predictiva.

[102] Como ejemplo, en el libro “Manual Práctico de Inteligencia Artificial en Entornos Sanitarios” establece que 80% del tiempo de desarrollo de un algoritmo tiene que ver con la obtención, la limpieza y el preprocesamiento de datos y que solo un 20% afecta al entrenamiento del algoritmo propiamente dicho

[103] Ver la Guía de Privacidad desde el Diseño de la AEPD

- Identificación y supresión, durante el proceso de entrenamiento, aquellas categorías de datos que no tienen una influencia significativa en el aprendizaje o en el resultado de la inferencia.
- Supresión de conclusiones no relevantes asociadas a información personal durante el proceso de entrenamiento, por ejemplo, en el caso de entrenamiento no-supervisado.
- Utilización de técnicas de verificación que requieran un menor número de datos, como la validación cruzada[104].
- Análisis y configuración de hiperparámetros[105] del algoritmo que pudieran tener influencia en la cantidad o extensión de datos tratados para minimizar estos.
- Utilización de modelos de aprendizaje federado en vez de centralizado[106].
- Aplicación de estratégicas de privacidad diferencial.
- Entrenamiento con datos cifrados utilizando técnicas homomórficas.
- Agregación de datos.
- Anonimización y seudonimización, no solo en la comunicación de datos, sino también en los datos de entrenamiento, posibles datos personales contenidos en el modelo[107] y en el tratamiento de la inferencia.

Extensión de las categorías de datos en la solución IA

En cada fase de un tratamiento, independientemente que incluya un componente IA o no, se necesitará procesar una extensión diferente del total datos personales relativos a un mismo interesado. Normalmente no será preciso que en cada fase se acceda a toda la extensión de datos personales disponibles[108], por lo que las estrategias de minimización de datos que se empleen en cada fase del tratamiento deberían ser distintas. También deberían serlo en cada una de las etapas del ciclo de vida de la solución IA, ya sea la etapa de entrenamiento, de inferencia o de evolución del modelo. Todo ello sin olvidar las limitaciones establecidas por la propia base jurídica[109].

En particular, un aspecto que hay que justificar a la hora de establecer la extensión de las categorías de datos a emplear en el componente IA es el de la utilización de variables proxy. Una variable proxy es un tipo de dato que, de por sí, parece no tener relación con el objeto del tratamiento, pero que podría tener una fuerte correlación con el valor inferido[110]. A la hora de plantear el tratamiento de variables proxy es necesario demostrar la validez de dicha correlación para legitimar su tratamiento.

Existen dos aspectos importantes que hay que analizar con relación a las variables proxy. El primero, es que su empleo no esté vinculado con algún sesgo en el modelo de razonamiento. Un ejemplo sería utilizar la nacionalidad del interesado como un elemento

[104] La validación cruzada supone la división de la base de datos en entrenamiento y testeo varias veces y con distinta selección de datos en cada subgrupo, comparando el resultado global del entrenamiento y prueba del algoritmo en cada iteración.
[105] Hiperparámetros son aquellos parámetros que permiten configurar el funcionamiento de un modelo específico de IA. Por ejemplo, en redes neuronales hiperparámetros serán el número de capas, el número de neuronas por capa, la tasa de aprendizaje, etc.
[106] Aprendizaje Federado o FL es un modelo de ML distribuido que permite el entrenamiento del modelo en los dispositivos finales (process over the edge). Es una aproximación "llevar el tratamiento a los datos en vez de los datos al tratamiento" y permite gestionar problemas de privacidad, propiedad y localización de los datos. https://arxiv.org/pdf/1902.01046.pdf
[107] Releasing the SVM Classifier with Privacy-Preservation https://ieeexplore.ieee.org/document/4781198
[108] Este principio está relacionado con el principio "need-to-know" que se emplea en seguridad, pero, en este caso, se interpreta desde el punto de vista de la privacidad. Por ejemplo, si el conjunto del tratamiento tiene una fase en la que se remite una comunicación al interesado, en dicha fase será necesario tratar los datos de contacto, pero no necesariamente tratar toda la información que se tiene del interesado. Incluso, para aquellos que ejecutan la comunicación, no será necesario acceder al contenido de la misma.
[109] La legitimación para tratar un dato en el marco de un tratamiento, por ejemplo, la dirección de entrega de un paquete, no se extiende a utilizar dicha dirección como dato de entrada al componente IA para perfilar la solvencia del cliente durante la fase de contratación.
[110] Variables proxy como el consumo de pizza en los centros de decisión de Estados Unidos o la ocupación del parking en los mismos, desveló la fecha precisa de la operación Tormenta del Desierto en 1991 y ha motivó cambio de los procedimientos de seguridad https://www.army.mil/article/2758/army_releases_new_opsec_regulation

para baremar su perfil de fraude o su situación social para evaluar su posibilidad de reincidencia en un delito. El segundo[111] aspecto es auditar que la correlación entre las variables proxy empleadas y el propósito de tratamiento permanece en el tiempo y en el ámbito de aplicación, ya que dicha relación es susceptible de cambiar y puede estar sujeta a fraude[112].

Extensión del conjunto de entrenamiento

Un aspecto específico del principio de minimización en soluciones IA es el relativo al tamaño de la población de la que se van a recoger los datos para entrenar los modelos, en particular cuando dicho tratamiento se base en el interés legítimo.

En dicho caso, el número de interesados afectados ha de justificarse en función:

- De la exactitud que la solución IA ha de alcanzar en sus inferencias para el tratamiento específico.
- Del factor de convergencia del algoritmo de entrenamiento hacia el valor de exactitud deseada en función del volumen de ejemplos proporcionados.
- De que el volumen de datos utilizado, bien por falta de disponibilidad de datos que cubran un espectro equilibrado de población (por edad, sexo, estatus, raza, cultura…), bien por la antigüedad de estos que reflejen contextos sociales obsoletos[113], introduzca sesgos en la inferencia.

El hecho de alimentar un modelo de aprendizaje de IA con datos sin ningún control ni análisis previo, además de no estar justificado, especialmente en el caso de que el tratamiento se base en el interés legítimo, puede hacer que la IA pierda precisión y se convierta en un multiplicador de sesgos. El conjunto de datos a utilizar ha de analizarse cuidadosamente para evitar dichos riesgos y legitimar su uso si, por ejemplo, el tratamiento se está basando en el interés legítimo.

Además, un exceso de datos de entrenamiento (sobreajuste o overfitting) puede provocar que el modelo esté muy ajustado a casos fácilmente reidentificables[114], siendo más vulnerable a ataques contra la privacidad (ver en el apartado “F. Seguridad” ataques por inversión del modelo).

Datos personales en la solución IA

En el caso de que se distribuya la IA como componente de un sistema y que, como parte de este, se incluyan datos de terceros, se ha de realizar una evaluación formal de qué datos personales de interesados de los comunicados podrían llegar a ser identificables.

Además, se han de aplicar medidas técnicas, organizativas o jurídicas para minimizar la identificación o la extensión de la identificación al mínimo número de datos posibles.

Por otro lado, durante la etapa de operación o inferencia, es conveniente utilizar estrategias de eliminación temprana de los datos del interesado en la solución IA, incluso cuando aplique la excepción doméstica.

[111] Está relacionado con el principio de exactitud.

[112] En el libro “Weapons of Math Destruction” referenciado en la bibliografía se describe el caso de utilizar la variable proxy “seguidores es Twitter” para evaluar la adecuación de los candidatos a un puesto de Experto en Social Media, en vez de evaluar cuidadosamente las campañas de marketing realizadas. Una vez que se conoce la variable proxy, es fácil engañar al sistema contratando por muy poco dinero un servicio que incrementa el número de seguidores en la cuenta Twitter del candidato

[113] Como utilizar información sobre las preferencias profesionales de las mujeres que incluya datos de hace muchos años, o fuera del contexto cultural, que creen una imagen obsoleta de la población actual.

[114] Además de otros problemas, como bajo rendimiento para una población que no está en el conjunto de entrenamiento.

F. SEGURIDAD

El RGPD establece en su artículo 32 que tanto el responsable, como el encargado, aplicarán medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado al riesgo para los derechos y libertades de los interesados. Estas medidas se adecuarán a los costes de aplicación, la naturaleza, el alcance, el contexto y los fines del tratamiento, así como a los riesgos de probabilidad y gravedad variables. Es decir, no hay una solución estándar para todos los tratamientos, y mucho menos para aquellos que incluyan un componente de IA. La solución tiene que establecerse mediante un análisis de riesgos que, desde el punto de vista de protección de datos, ha de ser relativo a los riesgos para los derechos y libertades de los interesados.

Amenazas específicas en componentes IA

El análisis de las medidas de seguridad comunes a cualquier sistema es igualmente necesario para proteger una solución de IA. Por ejemplo, deben considerarse aquellas que se derivan de que los componentes sean desarrollados por terceros o de las comunicaciones de datos a terceras partes que puedan producirse. Además, hay que tener en cuenta la necesidad de implementar medidas concretas dirigidas a prevenir ataques que son específicos a este tipo de aplicaciones.

Existen tipologías estudiadas de ataque y defensa a componentes de IA[115]. Entre las medidas de seguridad, se recomienda prestar especial atención a aquellas que gestionen estos tipos de amenazas:

- Acceso y manipulación del conjunto de datos de entrenamiento, previo a la configuración del modelo, por ejemplo, mediante técnicas de envenenamiento con patrones adversos.
- Inclusión de troyanos[116] y puertas traseras[117] durante el proceso de desarrollo de la IA, bien en el propio código o en las herramientas de desarrollo[118].
- Manipulación de la API de usuario para realizar accesos al modelo, tanto a nivel de caja negra como de caja blanca, para la manipulación de parámetros del modelo, filtrado del modelo a terceros, ataques a la integridad o disponibilidad de las inferencias[119].
- Ataques por "adversarial machine learning"[120] por lo que sería necesario un análisis de la robustez[121] y control de la alimentación con datos al modelo.
- Ataques por imitación de patrones que se conoce serán admitidos por el sistema[122].

[115] A Survey on Security Threats and Defensive Techniques of Machine, Qiang Liu et al., IEEE Access, ISSN:2169-3536, febrero 2018
[116] Trojans in IA, IARPA https://www.iarpa.gov/index.php?option=com_content&view=article&id=1150&Itemid=448
[117] Ejemplo de manipulación de un algoritmo de reconocimiento de imagen https://www.bleepingcomputer.com/news/security/ai-training-algorithms-susceptible-to-backdoors-manipulation/ también Backdoor Embedding in Convolutional Neural Network Models via Invisible Perturbation https://arxiv.org/pdf/1808.10307.pdf
[118] Vulnerabilidad en una librería de desarrollo de modelos ML: https://cyware.com/news/critical-vulnerability-in-numpy-could-allow-attackers-to-perform-remote-code-execution-33117832
[119] Con el propósito de incrementar la tasa de falsos positivos o falsos negativos.
[120] Técnica de ataque que consiste en alimentar la IA con datos de ejemplo, que desde el punto de vista de la percepción humana pueden resultar indistinguibles de datos normales, pero que incluyen pequeñas perturbaciones que fuerzan a la IA a realizar inferencias erróneas.
[121] Exploring the Landscape of Spatial Robustness, Logan Engstrom∗ MIT https://arxiv.org/pdf/1712.02779.pdf
[122] Orientados a aplicaciones como reconocimiento facial o detecciones de intrusiones y relacionados con técnicas de "adversarial machine learning" en muchas ocasionas.

- Reidentificación de los datos personales incluidos en el modelo (inferencia de pertenencia[123] o inversión del modelo[124]) por parte de usuarios internos[125] y externos.
- Fraude o engaño a la IA por parte de los interesados, especialmente en casos que puedan suponer un perjuicio para otros interesados[126], lo que implica la necesidad de realizar un análisis de la robustez ante dichas actuaciones y la realización de auditorías.
- Filtrado a terceros de resultados de perfilado o decisiones inferidas por la IA (también relacionado con las API's de usuario).
- Filtrado o acceso a los logs resultado de las inferencias generadas en la interacción con los interesados.

Logs o registros de actividad

La existencia de ficheros de log o registro de actividad, la realización de auditorías (ya sean automáticas o manuales) y la certificación del proceso, son una parte consustancial a las estrategias de "accountability" o responsabilidad proactiva. También se derivan de requisitos legales especialmente establecidos en la normativa sectorial.

Los ficheros de log serán necesarios para sustentar los procesos de auditoria y los mecanismos de seguridad. Con relación a protección de datos, dichos ficheros de log tendrán que proporcionar evidencias para:

- Determinar quién y bajo qué circunstancias accede a los datos personales que pudieran estar incluidos en el modelo.
- Proporcionar trazabilidad en cuanto a la actualización de los modelos de inferencia, las comunicaciones del API del usuario con el modelo y la detección de intentos de abuso o intrusión.
- Proporcionar trazabilidad para permitir la gobernanza en la comunicación de datos entre todos los intervinientes en la solución IA en cuanto a las obligaciones que se derivan del Considerando 66 del RGPD.
- Proporcionar seguimiento de los parámetros de calidad de la inferencia cuando la IA se utilice para la toma de decisiones o procesos de ayuda a la toma de decisión.

El interés legítimo del responsable, como base legitimadora para el tratamiento de los datos personales de los ficheros de registros de actividad con propósitos de seguridad, se desarrolla en el Considerando 49[127] del RGPD, en "*la medida estrictamente necesaria y proporcionada para garantizar la seguridad de la red y de la información*". En otros casos, la normativa sectorial establece obligaciones sobre la conservación y tratamiento de los registros de actividad. Es el caso de la Ley 10/2010 de prevención del blanqueo de capitales

[123] Cuando se puede determinar si un determinado individuo, sus datos, está o no formando parte del modelo de entrenamiento.
[124] El ataque por inversión del modelo en ML se produce cuando el atacante tiene acceso a ciertos datos personales del interesado incluidos en el modelo IA y puede inferir información personal adicional de dichos individuos analizando las entradas y salidas del modelo.
[125] En particular cuando los modelos de IA son adquiridos al desarrollador por terceros.
[126] Un ejemplo clásico de fraude en el de análisis mediante IA de CVs es el de escribir méritos inexistentes en un color de letra igual que el fondo del documento, imposible de leer para un humano, pero sí para una máquina, de forma que se puede engañar al sistema de selección de candidatos en perjuicio de candidatos honestos. Este sistema ya se utilizó para engañar al buscador Google y que este indexase páginas por palabras claves que no eran visibles al usuario, y no tenían relación con el contenido real de la página.
[127] Cons. 49 - Constituye un interés legítimo del responsable del tratamiento interesado el tratamiento de datos personales en la medida estrictamente necesaria y proporcionada para garantizar la seguridad de la red y de la información, es decir la capacidad de una red o de un sistema información de resistir, en un nivel determinado de confianza, a acontecimientos accidentales o acciones ilícitas o malintencionadas que comprometan la disponibilidad, autenticidad, integridad y confidencialidad de los datos personales conservados o transmitidos, y la seguridad de los servicios conexos ofrecidos por, o accesibles a través de, estos sistemas y redes, por parte de autoridades públicas, equipos de respuesta a emergencias informáticas (CERT), equipos de respuesta a incidentes de seguridad informática (CSIRT), proveedores de redes y servicios de comunicaciones electrónicas y proveedores de tecnologías y servicios de seguridad. En lo anterior cabría incluir, por ejemplo, impedir el acceso no autorizado a las redes de comunicaciones electrónicas y la distribución malintencionada de códigos, y frenar ataques de «denegación de servicio» y daños a los sistemas informáticos y de comunicaciones electrónicas

y de la financiación del terrorismo[128]. En ese supuesto, la base jurídica para realizar el tratamiento sería el cumplimiento de una obligación legal aplicable al responsable del tratamiento. Asimismo, otras bases jurídicas podrían emplearse para legitimar el tratamiento de datos en los registros de log.

El responsable deberá ser consciente de las obligaciones y límites establecidos en la norma sectorial, porque estas bases jurídicas no habilitan para el tratamiento de los datos personales contenidos en el fichero de log con otros fines distintos como podría ser la evaluación del rendimiento o la evolución del sistema IA. El responsable ha de asegurarse que se implementan garantías para evitar el acceso y la explotación de dicho registro con propósitos para los que se carece de base jurídica.

Los desarrolladores del componente IA, cuando estén utilizando soluciones basadas en ML y con propósito de documentar y de cumplir con el principio de responsabilidad proactiva deben implementar otros registros, como aquellos que permiten realizar la trazabilidad sobre la procedencia de los datos de entrenamiento y validación, así como registros de los análisis que se han realizado sobre la validez de dichos datos y sus resultados.

G. EVALUACIÓN DE LA PROPORCIONALIDAD Y NECESIDAD DE DICHOS TRATAMIENTO

El artículo 35.7.b del RGPD establece que, a la hora de realizar una EIPD, se ha de llevar a cabo una evaluación de la necesidad y proporcionalidad del tratamiento respecto a su finalidad.

El empleo de soluciones basadas en IA sitúa a responsables y encargados en un escenario en el que el tratamiento, por sus características y, en particular, por la tecnología elegida, puede conllevar un alto nivel de riesgo. Por lo tanto, debería valorarse si el objeto del tratamiento no puede ser conseguido utilizando otro tipo de solución que alcance la misma funcionalidad, con un margen de rendimiento aceptable y un nivel de riesgo menor.

Es decir, la disponibilidad o novedad de una tecnología no justifica, por sí misma, su utilización, sino que debe ser objeto de ponderación, realizando un análisis de si el tratamiento, en la forma en que se plantea, es equilibrado porque se derivan más beneficios y ventajas concretas para el interés general y la sociedad en su conjunto que perjuicios, entendidos estos como los riesgos sobre los derechos y libertades de los sujetos cuyos datos son objeto de tratamiento.

[128] Artículo 25. Conservación de documentos.
1. Los sujetos obligados conservarán durante un período de diez años la documentación en que se formalice el cumplimiento de las obligaciones establecidas en la presente ley, procediendo tras el mismo a su eliminación. Transcurridos cinco años desde la terminación de la relación de negocios o la ejecución de la operación ocasional, la documentación conservada únicamente será accesible por los órganos de control interno del sujeto obligado, con inclusión de las unidades técnicas de prevención, y, en su caso, aquellos encargados de su defensa legal.
En particular, los sujetos obligados conservarán para su uso en toda investigación o análisis, en materia de posibles casos de blanqueo de capitales o de financiación del terrorismo, por parte del Servicio Ejecutivo de la Comisión o de cualquier otra autoridad legalmente competente:
a) Copia de los documentos exigibles en aplicación de las medidas de diligencia debida, durante un periodo de diez años desde la terminación de la relación de negocios o la ejecución de la operación.
b) Original o copia con fuerza probatoria de los documentos o registros que acrediten adecuadamente las operaciones, los intervinientes en las mismas y las relaciones de negocio, durante un periodo de diez años desde la ejecución de la operación o la terminación de la relación de negocios.
2. Los sujetos obligados, con las excepciones que se determinen reglamentariamente, almacenarán las copias de los documentos de identificación a que se refiere el artículo 3.2 en soportes ópticos, magnéticos o electrónicos que garanticen su integridad, la correcta lectura de los datos, la imposibilidad de manipulación y su adecuada conservación y localización.
En todo caso, el sistema de archivo de los sujetos obligados deberá asegurar la adecuada gestión y disponibilidad de la documentación, tanto a efectos de control interno, como de atención en tiempo y forma a los requerimientos de las autoridades

La norma ISO-3100 “Gestión del Riesgo. Principios y Directrices” expone en el apartado 5.5 “Tratamiento del Riesgo” las condiciones generales para la gestión del riesgo en cualquier tipo de ámbito:

> *“LA SELECCIÓN DE LA OPCIÓN MÁS APROPIADA DE TRATAMIENTO DEL RIESGO IMPLICA OBTENER UNA COMPENSACIÓN DE LOS COSTES Y LOS ESFUERZOS DE IMPLEMENTACIÓN EN FUNCIÓN DE LAS VENTAJAS QUE SE OBTENGAN, TENIENDO EN CUENTA LOS REQUISITOS LEGALES, REGLAMENTARIOS Y DE OTRO TIPO, TALES COMO LA RESPONSABILIDAD SOCIAL Y LA PROTECCIÓN DEL ENTORNO NATURAL. LAS DECISIONES TAMBIÉN SE DEBERÍAN TOMAR TENIENDO EN CUENTA LOS RIESGOS CUYO TRATAMIENTO NO ES JUSTIFICABLE EN EL PLANO ECONÓMICO, POR EJEMPLO, RIESGOS SEVEROS (CONSECUENCIAS ALTAMENTE NEGATIVAS) PERO RAROS (BAJA PROBABILIDAD).”*

En el ámbito de la protección de datos de carácter personal, los riesgos que se han de tener en cuenta son los relativos a los derechos y libertades de los interesados. Si los riesgos los está asumiendo la sociedad, las ventajas obtenidas no sólo han de ser evaluadas desde la perspectiva de la entidad, sino también desde el punto de vista social en el ámbito en el que se despliegue el tratamiento.

En particular, en el caso de tratamientos que hagan uso de soluciones IA para la toma de decisiones o para la ayuda a la toma de dichas decisiones, se recomienda que en la EIPD se valore llevar a cabo un análisis comparativo entre el rendimiento obtenido por un operador humano cualificado frente a los resultados arrojados por modelos capaces de predecir escenarios o tomar acciones de manera automática[129]. En ese caso, deben tenerse en cuenta las condiciones reales de entrada de los datos así como el contexto en el que se despliega el tratamiento. Además, dicho análisis tendría que cubrir tanto los aspectos estrictos de la toma de decisión como los beneficios colaterales para los interesados.

H. AUDITORIA

El proceso de auditoría sobre un tratamiento que incluya un componente IA puede tener distinta extensión: auditoría sobre el proceso de desarrollo del componente, sobre la distribución del modelo, sobre aspectos concretos del tratamiento, de la operación, de la seguridad y robustez ante ataques al modelo, etc. Además, la auditoría puede ser realizada de forma interna o externa, y tener tanto un propósito de control como ser una herramienta de transparencia.

Como establece el principio de responsabilidad proactiva, las garantías establecidas para gestionar el riesgo han de estar documentas y recoger la información suficiente para permitir, de forma satisfactoria y demostrable, acreditar las acciones tomadas. Esta documentación ha de permitir la trazabilidad de las decisiones y comprobaciones realizadas siguiendo los principios de minimización antes señalados. No sólo hay que auditar el tratamiento, sino que este ha de ser auditable a lo largo de su ciclo de vida, incluyendo su retirada.

Es necesario realizar la auditoría para determinar la adecuación a las exigencias del RGPD y comprobar la validez del tratamiento basado en soluciones de IA. El proceso de auditoría, para ser efectivo, ha de realizarse en las mismas condiciones que un entorno real de explotación, en particular, para evaluar:

- La existencia de un proceso de análisis, desarrollo y/o de implementación documentado incluyendo, cuando proceda, las oportunas evidencias de trazabilidad.
- La existencia o no de datos personales, elaboración de perfiles o decisiones automáticas sobre interesados sin intervención humana, así como el análisis de la eficacia de los métodos de anonimización y seudonimización.

[129] En particular en aplicaciones donde la complejidad del tratamiento tiene múltiples aspectos y elementos colaterales que no pueden simplificarse. Una reflexion en este aspect se puede encontrar en. Why we cannot trust artificial intelligence in medicine, Matthew DeCamp, Jon C Tilburt, The Lancet, correspondence, volume 1, issue 8, december 01, 2019

- Análisis de la existencia y legitimación del tratamiento de categorías especiales de datos, en particular en la información inferida.
- La base jurídica para el tratamiento y la identificación de responsabilidades.
- En particular, cuando la base jurídica es el interés legítimo, evaluación del balance entre los distintos intereses e impactos sobre los derechos y libertades en función de las garantías adoptadas.
- La información y la efectividad de los mecanismos de transparencia implementados.
- La aplicación del principio de responsabilidad proactiva y la gestión del riesgo para los derechos y libertades de los interesados y en particular, si se ha evaluado la obligación o necesidad de la ejecución de EIPDs y, en caso afirmativo, sus resultados.
- Con relación a lo anterior, la aplicación de medidas de protección de datos por defecto y desde el diseño, entre otras:
    - El análisis previo de las necesidades de tratamiento de datos personales, en cantidad y extensión, en las distintas etapas siguiendo criterios de minimización.
    - El análisis de la exactitud, fidelidad, calidad y sesgos de los datos utilizados o capturados para el desarrollo o la operación del componente IA, así como los métodos de depuración de datos utilizados.
    - La comprobación y realización de procesos de pruebas y validación de la precisión, exactitud, convergencia, consistencia, predictibilidad y cualquier otra métrica de la bondad de los algoritmos empleados, perfilados e inferencias. Además, la verificación de que estos parámetros cumplen los requisitos necesarios para el tratamiento.
- La adecuación de las medidas de seguridad para evitar riesgos a la privacidad.
- La capacitación y formación del personal del responsable del tratamiento vinculado, cuando proceda, con el desarrollo o la explotación del componente IA, en este último caso con especial atención a la interpretación correcta de las inferencias.
- La necesidad y, si procede, capacidad del DPD.
- La incorporación de mecanismos que garanticen la atención a los derechos de los interesados, en particular la supresión de oficio de datos personales, y prestando especial cuidado a los derechos de menores.
- El cumplimiento de las limitaciones sobre las decisiones automáticas sin intervención humana, la evaluación, en su caso, de la calidad de la intervención humana y los mecanismos de supervisión adoptados. En particular, cuando la base jurídica sea el consentimiento explícito, identificación de las garantías adoptadas para determinar si éste es libre.
- La aplicación de alguna de las garantías que se establecen en el Capítulo V del RGPD en el caso de que existan transferencias internacionales de datos.
- En general, el cumplimiento de los requerimientos y obligaciones del RGPD y, en particular, los expuestos en este documento.

Como se ha establecido anteriormente, la solución IA estará integrada en un tratamiento específico, con unas características singulares y en un entorno de explotación determinado. Una auditoría de la solución IA aislada, sin tener en cuenta el contexto y el entorno, estará incompleta y ofrecerá resultados parciales que no serán realistas[130].

Un aspecto crítico de la auditoría es garantizar que la solución IA se está empleando para el propósito para el que fue diseñada, con especial detalle a su utilización por los operadores

[130] Por ejemplo, la utilización de curvas ROC para demostrar el rendimiento de los sistemas de reconocimiento facial han de elaboradas en entornos que reflejen las condiciones reales de operación.

del sistema. Además, cuando se trata de un componente adquirido a un tercero, hay que determinar qué otras funciones colaterales está realizando dicho componente y si pueden tener repercusiones normativas[131].

La utilización de soluciones o herramientas de auditoría automática en tiempo real es recomendable en sistemas de toma de decisiones automatizadas para asegurar que los resultados son coherentes y precisos, además de para soportar la posibilidad de que las decisiones erróneas sean abortadas o canceladas antes de que se produzcan consecuencias irreversibles.

[131] Por ejemplo, el componente recaptcha de Google realiza también funciones analíticas: https://developers.google.com/recaptcha/docs/analytics

V. TRANSFERENCIAS INTERNACIONALES

El desarrollo o despliegue de un componente de IA basado en servicios en la Nube, la comunicación de los datos de los usuarios a terceros para evolucionar el modelo de IA, o la distribución de componente de IA en el caso de que existan datos personales inherentes al modelo, pueden implicar flujos transfronterizos de datos a terceros países. No tienen consideración de transferencia internacional de datos los flujos de datos que se producen dentro del marco del Espacio Común Europeo (los Estados de la Unión Europea más Islandia, Noruega y Liechtenstein).

Dichas transferencias han de aplicar las garantías que se establecen en el Capítulo V del RGPD "*Transferencias de datos personales a terceros países u organizaciones internacionales*". Especialmente importante es establecer mecanismos para permitir que las contrataciones que se realicen en este contexto de transferencias internacionales se gestionen con fluidez, asegurando al mismo tiempo que el cliente responsable tiene información suficiente sobre los contratistas, o potenciales contratistas, y mantiene la capacidad de tomar decisiones. Cuando existen transferencias internacionales hay que informar a los interesados en los términos del art. 13 y 14 del RGPD e incluirlo en el registro de actividades de tratamiento.

VI. CONCLUSIONES

La puesta en el mercado, para entidades y consumidores en general, de tratamientos que incluyan soluciones basadas en tecnologías disruptivas, como las basadas en componentes IA, exige que se implementen garantías de calidad y seguridad[132]. La disponibilidad de una tecnología o su novedad no es razón suficiente para comercializar productos que no cumplan con un nivel adecuado de calidad de servicio, especialmente cuando estos requisitos están establecidos en una norma. Por otro lado, los investigadores y la industria basada en IA necesitan guías y ayudas que les sirvan de apoyo en el cumplimiento normativo y les aporte seguridad jurídica en sus proyectos, productos y servicios.

Con relación a la protección de datos personales, el cumplimiento de lo establecido en el RGPD exige cierto nivel de madurez a las soluciones IA que permita determinar de forma objetiva la adecuación de los tratamientos y la existencia de avales y medidas para gestionar sus riesgos.

Para estos riesgos, la tecnología en general[133], y en particular las tecnologías de IA, puede tener un efecto multiplicador de las carencias éticas que ya están presentes en la sociedad o aquellas que se arrastran desde el pasado y que están registradas en datos históricos. La aplicación del RGPD y de las garantías que incorpora permite minimizar dichos riesgos.

El empleo de estrategias de transparencia, gestión del riesgo y mecanismos de auditoria y certificación no sólo permitirán el cumplimiento de lo establecido en el RGPD, sino que mejorarán la confianza de los usuarios en los productos y servicios basados en IA, además de abrir un nuevo mercado en este sector de actividad: ingenieros de privacidad, auditores, esquemas de certificación, profesionales acreditados, etc. Estas nuevas oportunidades de desarrollo se extienden también a la creación de esquemas de portabilidad.

Es más, las aplicaciones de IA pueden suponer un gran apoyo para proteger la privacidad e implementar mecanismos que aseguren la protección de datos[134].

Este documento pretende ser una mera introducción a la adecuación de los tratamientos que incluyan componentes de IA y no cubre todas las posibilidades y riesgos que se pueden derivar del empleo de soluciones en IA en tratamientos de datos personales[135].

Para finalizar, hay que subrayar que uno de los principales problemas de las soluciones de IA no es la IA en sí, sino cómo van a usar las personas la tecnología IA y los nuevos sesgos psicológicos que se derivan de su empleo. En particular, es necesario prestar especial atención a atribuir responsabilidades a componentes IA sin supervisión y sin adoptar una posición crítica. La delegación de la toma de decisiones a máquinas no es nueva; ya ocurre con los algoritmos deterministas, pero el sesgo de atribuir una autoridad o peso superior a un resultado inferido por una solución de IA puede hacer incrementar los riesgos derivados de esta delegación de responsabilidad.

[132] Seguridad entendida como “safety”, es decir, que no produce un daño o perjuicio.
[133] Como en el caso de situaciones de bulling, acoso o fake news.
[134] Como son las aplicaciones IA para depurar bases de datos o para analizar la seguridad de los sistemas.
[135] Hay muchas otras cuestiones que quedan pendientes, como son los casos de hibridación hombre-máquina, el problema de la privacidad grupal, la demutualización, o incluso hasta qué punto la decisión tomada por una IA en aplicaciones militares entra dentro de la categoría de decisiones automáticas con efectos jurídicos

VII. REFERENCIAS

Las referencias que se enumeran a continuación están ordenadas en función de la relevancia de estas a la hora de elaborar el presente documento:

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales

Big data, artificial, intelligence, machine learning and data protection. ICO information commissioner's office, septiembre de 2017

How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence, CNIL, diciembre 2017

Draft Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, European Commission, diciembre 2018, Brussels

Artificial Intelligence for Europe. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions -, Comisión Europea, abril 2018, Bruselas.

Discrimination, artificial intelligence, and algorithmic decision-making, Prof. Frederik Zuiderveen Borgesius, Consejo de Europa 2018

Robustness and Explainability of Artificial Intelligence. From technical to policy solutions. Comisión Europea. 2020

Estrategia Española de I+D+I en Inteligencia Artificial, Ministerio de Ciencia, Innovación y Universidades, 2019

Automating Society, Taking Stock of Automated Decision-Making in the EU, AW AlgorithmWatch gGmbH, enero 2019, Berlín

Opinion 7/2015 Meeting the challenges of big data. European Data Protection Supervisor. EDPS, 19 November 2015.

Manual Práctico de Inteligencia Artificial en Entornos Sanitarios, Beunza Nuin J.J., Puertas Sanz E., Emilia Condés Moreno E., Elsevier, enero 2020

Recomendaciones para el tratamiento de datos personales en la inteligencia artificial. Texto aprobado por las Entidades integrantes de la Red Iberoamericana de Protección de Datos en la sesión del 21 de junio de 2019, en la ciudad de Naucalpan de Juárez, México.

Artificial Intelligence and Data Protection in Tension, Centre for Information Policy Leadership, Octubre 2018

AI and Data Protection – Balancing tensions – Slaughter and May, PLC Magazine, agosto 2019

Artificial Intelligence in Finance, Instituto Alan Turing, Hanken School of Economics, abril 2019, Finlandia

Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System, Partnership on AI (PAI), abril 2019 https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/

Human-AI Collaboration Framework & Case Studies, Partnership on AI (PAI), septiembre 2019, https://www.partnershiponai.org/human-ai-collaboration-framework-case-studies/

AI Index 2018 Annual Report, Zoe Bauer et al., AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, December 2018

Artificial Intelligence and the Future of Humans, J. Anderson et al., Pew Research Center, diciembre 2018

Serie relativa a Inteligencia Artificial del blog del ICO https://ico.org.uk/about-the-ico/what-we-do/tech-and-innovation/blogs/

Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures Roman V. Yampolskiy, University of Louisville 2018

Why we cannot trust artificial intelligence in medicine, Matthew DeCamp, Jon C Tilburt, The Lancet, correspondence, volume 1, issue 8, december 01, 2019

Exploring the landscape of spatial robustness, Logan Engstrom, MIT https://arxiv.org/pdf/1712.02779.pdf

Towards federated learning at scale: system design Keith, Bonawitz et al. https://arxiv.org/pdf/1902.01046.pdf

What's your ML Test Score? A rubric for ML production systems, Eric Breck, 2018, Google Inc. https://www.eecs.tufts.edu/~dsculley/papers/ml_test_score.pdf

A Survey on Security Threats and Defensive Techniques of Machine, Qiang Liu et al., IEEE Access, ISSN: 2169-3536, febrero 2018

Dictamen 05/2014 sobre técnicas de anonimización, Grupo del Artículo 29, 2014

Dictamen 06/2014 sobre el concepto de interés legítimo del responsable del tratamiento de los datos en virtud del artículo 7 de la Directiva 95/46/CE, Grupo del Artículo 29, abril 2014

Directrices sobre el derecho a la portabilidad de los datos Grupo del Artículo 29. Adoptadas el 13 de diciembre de 2016 Revisadas por última vez y adoptadas el 5 de abril de 2017

Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679. Grupo del Artículo 29. Adoptadas el 3 de octubre de 2017 Revisadas por última vez y adoptadas el 6 de febrero de 2018

Directrices sobre los delegados de protección de datos (DPD), Grupo del Artículo 29. Adoptadas el 13 de diciembre de 2016 Revisadas por última vez y adoptadas el 5 de abril de 2017

Guía de Privacidad desde el Diseño, AEPD, octubre 2019

Guía práctica para las evaluaciones de impacto en la protección de datos personales, AEPD, octubre 2018

Orientaciones y garantías en los procesos de anonimización de datos personales, AEPD, 2016

Listas de tipos de tratamientos de datos que requieren EIPD (art 35.4) AEPD, 2019

Modelo de informe de Evaluación de Impacto en la Protección de Datos para Administraciones Públicas, AEPD, julio 2019

ISO-3100 "Gestión del Riesgo. Principios y Directrices"

VIII. ANEXO: SERVICIOS ACTUALES BASADOS EN IA

La lista que se presenta en este anexo no pretende ser exhaustiva, sino un ejemplo para ilustrar la extensión de los servicios que actualmente se están prestando basándose en IA:

- Servicios de Internet
  - Captchas, chatbots, detección de fraude, personalización de anuncios.
- Recursos humanos
  - Selección de candidatos.
- Servicios financieros
  - Predicción de hipotecas en base al análisis del perfil del cliente, monitorización de transacciones para detectar actividades fraudulentas basándose en los hábitos de consumo, inversión financiera automática.
- Salud y Sanidad
  - Diagnóstico basado en el análisis de imágenes, predicción de tasas de readmisión de pacientes en base al análisis de los datos, mapas sanitarios, análisis de salud mental, prevención de suicidios, chatbots de salud mental, predicción de riesgo basado en parámetros analíticos, diagnóstico por análisis de muestra patológica, procesamiento del lenguaje natural de historias clínicas, análisis genético, electrodiagnóstico, desarrollo de vacunas y medicamentos.
- Comercio y comunicación:
  - Recomendaciones de productos basándose en el perfil del cliente y en el análisis de sus compras, maximizar el alcance de productos y servicios a un grupo de clientes, agentes de viaje virtuales, monitorización de redes sociales.
- Servicios públicos y suministros:
  - Contadores inteligentes y predicción de la demanda de consumo de los clientes, estimación del coste de determinados servicios de mantenimiento, asignación de tratamientos en el sistema público de sanidad, tratamiento automático de multas, soporte a la decisión en administración de justicia.
- Transporte:
  - Vehículos autónomos, semáforos inteligentes, optimización de las rutas y horarios de los servicios públicos de transporte.
- Educación:
  - Contenido y formación personalizada a las necesidades del alumnado, corrección de exámenes, detección de plagio o fraude en trabajos, tutorización automática, detección de estudiantes anómalos.
- Seguridad:
  - Reconocimiento facial, huellas dactilares, detección de comportamiento, control de fronteras, análisis de indicio de engaño, análisis de registros de actividad, detección de intrusiones, análisis de comunicaciones.
- Hogar:

- Asistentes inteligentes, espejos inteligentes, electrodomésticos, seguridad.

- Otros
  - Herramientas de dibujo, ayudas a la creación artística, optimización de programas de entrenamiento deportivo.