

Exactitud, idoneidad y calidad de los datos en tratamientos de datos personales con Inteligencia Artificial

RESUMEN EJECUTIVO

Este documento formula una serie de reflexiones acerca de la relación entre la calidad de los datos y el principio de exactitud previsto en el Reglamento General de Protección de Datos (RGPD) en el marco de los tratamientos de datos personales, poniendo de manifiesto que ambos conceptos deben interpretarse en función de la finalidad perseguida y de la idoneidad del tratamiento. Aunque estrechamente relacionados, los requisitos de calidad de los datos y el principio de exactitud del RGPD no son conceptos equivalentes. La calidad de los datos constituye un concepto más amplio, aplicable tanto a datos personales como no personales y debe dar respuesta a requisitos más allá del RGPD.

La interpretación del principio de exactitud del RGPD no debería limitarse exclusivamente a la veracidad o a la actualización de los datos. Dependiendo de la finalidad del tratamiento, la adecuación de los datos puede requerir también otras propiedades, tales como la precisión, el nivel de granularidad, la frecuencia de actualización o cualquier otra característica cuya ausencia pueda comprometer el cumplimiento de la finalidad perseguida. La veracidad y la actualidad de los datos deben exigirse cuando resulten necesarias para dicha finalidad, especialmente cuando su falta pueda incidir en los derechos y libertades de las personas físicas, sin que ello implique que constituyan requisitos exigibles para todo tratamiento.

Asimismo, cuando un tratamiento de datos personales incorpore datos no personales, estos también deberán satisfacer los requisitos de calidad que resulten adecuados para la finalidad perseguida, en la medida en que puedan influir en la idoneidad del tratamiento o en aquellos resultados susceptibles de producir efectos sobre los derechos y libertades de las personas físicas.

La evaluación de la calidad no puede limitarse a los datos de entrada. Resulta igualmente necesario establecer métricas objetivas sobre la calidad de los resultados generados por el tratamiento, ya que, sin ellas, no es posible verificar ni demostrar que la finalidad perseguida se alcanza de forma eficaz y que el tratamiento mantiene las condiciones de idoneidad exigibles.

Esta relación entre calidad e idoneidad adquiere una relevancia especial en los tratamientos destinados al desarrollo o evolución de sistemas y modelos de inteligencia artificial. Los conjuntos de datos que no reúnen los requisitos de calidad para dicho desarrollo no pueden considerarse necesarios para el tratamiento. En consecuencia, el acceso a tales conjuntos de datos únicamente podría justificarse en la medida estrictamente necesaria para evaluar objetivamente su calidad y determinar su eventual aptitud para la finalidad prevista.

Los mecanismos de gobierno y los procesos de gestión de datos han de coordinar tanto las obligaciones derivadas de los tratamientos de datos personales como las exigencias propias del desarrollo, adquisición, despliegue y evolución de sistemas de inteligencia artificial, garantizando la coherencia de los criterios de calidad y de cumplimiento normativo a lo largo de todo el ciclo de vida.

Finalmente, la definición, evaluación y revisión de los requisitos de calidad de los datos exige un enfoque multidisciplinar. Para ello resulta esencial contar, al menos, con la participación conjunta de especialistas en ciencia de datos y en protección de datos, que se guíen por criterios objetivos y evidencia verificable durante todo el ciclo de vida de los tratamientos.

Palabras clave: inteligencia artificial, exactitud, minimización, finalidad, idoneidad, calidad de datos, conjuntos de datos, gobernanza.

CONTENIDO

I.	INTRODUCCIÓN	4
II.	CALIDAD DE LOS DATOS Y EXACTITUD-RGPD	4
A.	Calidad de los datos	4
B.	Exactitud y minimización en el RGPD	5
C.	Características de calidad de los datos en los estándares	6
D.	Características de calidad de los datos más allá del estándar	7
E.	Calidad del dato individual vs calidad del conjunto de datos	8
F.	Calidad de los datos vs veracidad y actualidad	9
G.	Alineación de exactitud-RGPD con calidad de los datos	10
III.	TRATAMIENTOS DE DATOS PERSONALES E IA	11
A.	Tipos de tratamientos de datos personales	12
B.	Calidad de los datos en tratamientos de datos personales que incluyen sistemas de IA	13
C.	Exigibilidad de calidad en los datos no personales dentro de tratamientos de datos personales	15
D.	Desarrollo de IA como tratamiento	15
E.	Determinación de los requisitos de calidad de los datos en el desarrollo de sistemas de IA	16
F.	Análisis de calidad de las fuentes de datos para IA	17
G.	Requisitos de calidad inadecuados	18
IV.	REFLEXIONES FINALES	18
V.	ANEXO: EVOLUCIÓN DEL CONCEPTO DE EXACTITUD EN LA NORMATIVA ESPAÑOLA	21

I. INTRODUCCIÓN

El concepto de calidad de los datos es esencial para la economía del dato cuando productos o servicios dependen, para su desarrollo, del acceso a conjuntos de datos. Además, tales conjuntos de datos precisan tener la calidad adecuada para alcanzar la finalidad prevista, y cumplir con distintos requisitos, en un contexto específico¹. Estos conjuntos deben contener datos representativos que ofrezcan la cantidad de información necesaria para garantizar el éxito de los procesos basados en datos, ya sean datos personales o no. Un caso particular sería el desarrollo de modelos y sistemas de IA² basados en técnicas de aprendizaje automático y/o que precisen conjuntos de datos para verificación, validación y prueba.

Por otro lado, los tratamientos de datos personales, que pueden incluir en sus operaciones uno o más sistemas de inteligencia artificial, deben ser conformes al RGPD^{3 4}, en particular con relación a los principios de exactitud y de minimización con relación a los datos personales. Asimismo, el resultado de un tratamiento de datos personales ha de cumplir la condición de idoneidad para cumplir su finalidad, y en cuanto el resultado sea relativo a personas identificadas o identificables, se ha de garantizar que se protegen sus derechos y libertades, en particular en el respeto de ambos principios. Un caso específico de tratamiento de datos personales es aquel en el que conjuntos de datos personales (y no personales) son necesarios para el desarrollo de modelos y sistemas de IA.

La presente nota técnica trata sobre la aplicación de dichos conceptos (exactitud, minimización, idoneidad y calidad de datos) en el marco de tratamientos a los que aplica el RGPD. Está orientada a responsables, encargados y DPD's, así como para aquellos implicados en el desarrollo, prueba y mantenimiento de productos o servicios que precisen de acceso a conjuntos de datos personales, como pueden ser muchos tipos de modelos y sistemas de IA.

II. CALIDAD DE LOS DATOS Y EXACTITUD-RGPD

A. CALIDAD DE LOS DATOS

La calidad de los datos se define en distintos estándares (por ejemplo, UNE 0079, ISO 25012, ISO 5259, ISO 8000) con relación a cualquier tipo de datos, ya sean datos personales o no, como el **grado** en que las características de los datos deben satisfacer **necesidades declaradas e implícitas** cuando se utilizan **en condiciones específicas**. La norma ISO 22989⁵, con relación a sistemas de IA, define el proceso de comprobación de la calidad del dato como aquel "*en el que los datos son examinados para determinar su completitud, sesgo y otros factores que afecten a su utilidad para un sistema de IA*". Estas definiciones sirven de orientación técnica, y no sustituyen ni modifican las exigencias jurídicas derivadas del RGPD.

Tales definiciones vinculan el concepto de calidad de los datos a la adecuación de estos para cumplir una **finalidad** ("*satisfacen necesidades declaradas e implícitas*" o "*utilidad para un sistema IA*") en un **contexto**⁶ determinado. Las "*necesidades declaradas e implícitas*" también podrían incluir el cumplimiento de obligaciones de distintas normativas, cumplimiento de estándares, compromisos contractuales y políticas internas de una entidad.

¹ Así es como se expresa en el RIA para los modelos y sistemas de su competencia en el considerando 12

² El término sistema o modelo de IA se utiliza en este texto con un significado más amplio que el utilizado en el RIA y en las guías de la Comisión Europea sobre la definición de sistema de IA, pudiendo incluir otras técnicas basadas en ciencia de datos, como las regresiones logísticas, optimizaciones matemáticas, métodos estadísticos, etc.

³ Deben garantizar y poder demostrar que son conformes, según el art.24 del RGPD

⁴ Aunque se utilice únicamente el término RGPD, se extiende a la competencia establecida en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

⁵ ISO 22989 Artificial intelligence concepts and terminology

⁶ En el RIA, se define en el considerando 12 el contexto en el que funciona el sistema IA

Estas necesidades específicas se representan a través de los llamados **requisitos de calidad de los datos**.

B. EXACTITUD Y MINIMIZACIÓN EN EL RGPD

Pese a que la calidad de los datos puede considerarse un principio clásico en el ámbito de la privacidad, lo cierto es que el RGPD no hace referencia al término “*calidad de los datos*”, aunque sí establece dos principios especialmente relacionados con él, como son el de minimización y el de exactitud en el artículo 5 apartado 1, letras c) y d):

“1. Los datos personales serán:

c) **adecuados, pertinentes y limitados a lo necesario en relación con los fines** para los que son tratados («**minimización de datos**»);

d) **exactos** y, si fuera **necesario, actualizados**; se adoptarán todas las medidas razonables para que se supriman o rectifiquen sin dilación los datos personales que sean inexactos **con respecto a los fines** para los que se tratan («**exactitud**»);”.

Si bien los requisitos de calidad de los datos no constituyen un principio del RGPD, su función sí es instrumental para facilitar la aplicación práctica de dichos principios. Ambos actúan de forma complementaria, coherente y autónoma para establecer la *banda de calidad* que han de tener los datos, exigiendo que sean correctos con relación a la finalidad y limitando el conjunto de datos a los necesarios. Tanto para los datos de entrada al tratamiento, como con relación a los resultados de salida del tratamiento.

OPERACIONES PARA LAS QUE SE PRECISA EL DATO

¿DÓNDE VIVES?		CONOCER EL PREFIJO TELEFÓNICO NACIONAL	DETERMINAR REDUCCIONES FISCALES	ENTREGA POSTAL
	España	Exacto	Inexacto	Inexacto
	ESPAÑA Madrid	Excesivo	Exacto	Inexacto
	ESPAÑA Madrid Gran Vía 320 1º	Excesivo	Excesivo	Exacto

Figura 1 Hay que encontrar el necesario balance entre exactitud y minimización para cumplir los fines del tratamiento

El cumplimiento de los principios de minimización de datos y de exactitud está **vinculado al de finalidad del tratamiento**⁷. En cuanto vinculado al cumplimiento de la finalidad, el significado del principio de exactitud de los datos en el RGPD (en adelante “exactitud-RGPD”⁹) tiene una extensión distinta, y no estrictamente vinculada a, por ejemplo, la veracidad y actualidad de cada uno de los datos. El término “exactitud-RGPD” está vinculado a que los datos sean adecuados para alcanzar la finalidad del tratamiento, es decir, conseguir que el tratamiento sea idóneo. Esto implica que los datos de entrada han de ser lo suficientemente exactos-RGPD para que el tratamiento consiga su finalidad y, si el resultado del tratamiento son datos personales, incluidas decisiones y

⁷ Excepto en el art.47 “Normas corporativas vinculantes” en el apartado 2 letra d) “Las normas corporativas vinculantes mencionadas en el apartado 1 especificarán, como mínimo, los siguientes elementos:...d) la aplicación de los principios generales en materia de protección de datos, en particular la limitación de la finalidad, la minimización de los datos, los periodos de conservación limitados, la calidad de los datos, la protección de los datos desde el diseño y por defecto, ...”.

⁸ En este caso, habría que considerar también el cumplimiento de principios como el de lealtad, sobre el que se podría fundamentar el tratamiento de datos para evitar, por ejemplo, sesgos, sin perjuicio de lo establecido en el artículo 9 del RGPD y el RIA.

⁹ El término exactitud con relación a los datos tiene un ámbito semántico distinto al del utilizado en los estándares antes mostrados y en otras normativas. No es el único caso. Términos que aparecen en el RGPD, como transparencia, interesado, tratamiento, consentimiento, responsable o riesgo entre otros, son empleados en otros contextos normativos con un significado muy distinto. Cuando se emplean en contextos en los que aplican distintos marcos regulatorios es importante añadir a los términos un “apellido” para determinar con precisión su significado.

perfilados, estos resultados han de respetar los derechos y libertades de las personas físicas, en particular, con relación a la exigencia de que los datos sean veraces.

El cumplimiento del principio de exactitud-RGPD puede implicar el satisfacer distintos requisitos “técnicos”, tales como una adecuada precisión, suficiente granularidad, una frecuencia de recogida adecuada, ausencia de sesgos, completitud de atributos, etc. Una interpretación extensiva del principio de exactitud-RGPD que fuera más allá del cumplimiento de la finalidad podría afectar a los derechos de los interesados, en la medida que habilitaría un tratamiento de datos que podrían ir más allá de las necesidades estrictas, contradiciendo el principio de minimización-RGPD.

Asimismo, dicha interpretación extensiva podría generar cargas desproporcionadas para los responsables y encargados y, en particular, para los desarrolladores/evolucionadores de sistemas de IA, al exigirles acceder y gestionar conjuntos de datos con unos niveles de, por ejemplo, veracidad, actualidad, precisión, etc., que podrían no ser necesarios. Incluso, esto podría ir en contra de la posibilidad de conseguir conjuntos de datos de calidad suficiente para desarrollar sistemas adecuados y, además, exigir a los responsables la atención de unos derechos más allá de lo establecido en el RGPD¹⁰. En todo caso, quedará a salvo la posibilidad de corrección de datos personales que sean inexactos en relación con la persona afectada en coherencia con el derecho de rectificación (artículo 16 RGPD).

C. CARACTERÍSTICAS DE CALIDAD DE LOS DATOS EN LOS ESTÁNDARES

Existen criterios para evaluar la calidad de los datos. Por ejemplo, en el estándar ISO 25012 se enumeran una serie de características para tener en cuenta: exactitud, completitud, consistencia, credibilidad, actualidad, accesibilidad, conformidad, confidencialidad, eficiencia, precisión, trazabilidad, comprensibilidad, disponibilidad, portabilidad y recuperabilidad. El estándar establece que su evaluación ha de realizarse en función del **contexto de uso específico**, por lo que no se trata de características absolutas sino relativas a dicho contexto.

Aunque en el listado de características del estándar se puede leer el término "exactitud", sería un error emparejar este término con el término exactitud-RGPD. Exactitud-ISO es el grado en el que los datos representan correctamente el verdadero valor de un atributo, de un concepto o de un evento en un contexto de uso específico. La exactitud-ISO podría ser una de las condiciones necesarias para alcanzar la exactitud-RGPD, es decir, para cumplir con la finalidad del tratamiento, pero no necesariamente. Además, para alcanzar la exactitud-RGPD podrían ser necesarias otras características del estándar, como la precisión, así como características que no se encuentran enumeradas en el estándar, como la ausencia de un sesgo.

Como todo estándar, el conjunto de características de calidad de los datos recoge la experiencia pasada y permite establecer un marco de referencia común que facilite la comparación de las evaluaciones obtenidas por distintos actores. Sin embargo, no resulta exigible el cumplimiento de todas las características, ni estas agotan los requisitos necesarios para garantizar la idoneidad respecto de cualquier finalidad. Por ello, no hay que cometer los siguientes errores al utilizar estos estándares:

- Aplicar “**calidad de checkbox**”, entendida como un repaso en abstracto del conjunto de datos sin tener en cuenta el contexto y la finalidad específicos, ni realizar un análisis basado en evidencias.
- Establecer que todas las características son igual de relevantes. Para un contexto y finalidad determinados, algunas de ellas podrían ser irrelevantes y, por tanto, no exigibles.

¹⁰ El art.16 del RGPD también vincula la rectificación a la finalidad del tratamiento.

- Suponer que las características a evaluar se **agotan en la lista del estándar**. El estándar no es exhaustivo, sino un valioso punto de partida que hay que adaptar a las circunstancias concretas de cada tratamiento.

Para no caer en estos errores, es necesario extraer la lista de características y sus requisitos para en función de los contextos y finalidades previstos del tratamiento.

D. CARACTERÍSTICAS DE CALIDAD DE LOS DATOS MÁS ALLÁ DEL ESTÁNDAR

Con relación al apartado anterior, y para el caso de desarrollo de modelos y sistemas de IA, puede ser necesario considerar más características con relación a la calidad de los datos, como por ejemplo¹¹:

- La eliminación de sesgos que comprometan la finalidad del tratamiento¹².
- Las propiedades estadísticas adecuadas para determinadas finalidades previstas.
- La representatividad del contexto con relación a las finalidades previstas.
- Que se reflejen los rasgos, características o elementos particulares del entorno geográfico, contextual, conductual o funcional específico en el que esté previsto que se utilice el sistema de IA.
- La frecuencia de recogida de datos, que puede ser relevante cuando se están realizando tratamientos basados en la geolocalización, en el consumo de energía, en la actividad en Internet, etc.
- El poder predictivo de las variables seleccionadas. Estas son aquellas con la capacidad de anticipar con precisión el valor de las variables resultado, al estar fuertemente relacionadas con ellas, permiten separar claramente entre clases y reducen la incertidumbre del resultado.
- La relevancia de muestras para el proceso de aprendizaje. Las muestras relevantes son aquellas que aportan información significativa y útil para adaptar el modelo en cada iteración de entrenamiento. Las técnicas de “*active-learning*” en aprendizaje automático están orientadas a seleccionar muestras relevantes, que dependerán del diseño del modelo, así como del grado de preentrenamiento que ya posee.

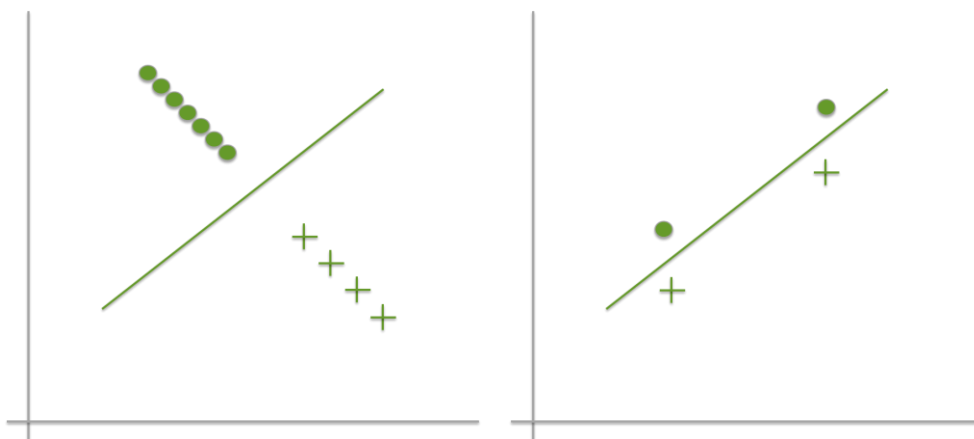


Figura 2 En un ejemplo, sencillo y evidente, de clasificación con machine learning en el que dos clases se han de dividir de forma lineal. El gráfico de la izquierda utiliza un conjunto de datos en el que la mayoría de las muestras no son relevantes para realizar una clasificación adecuada, mientras que el de la derecha es un conjunto de datos muy relevantes para obtener dicha clasificación lineal.

¹¹ Están enumeradas en el considerando 67 y artículo 10 del RIA en los apartados del 1 al 4.

¹² Los sesgos no son en todos los casos negativos, sino que, en algunos de ellos, tanto los sesgos en los conjuntos de datos utilizados en aprendizaje automático pueden ser necesarios en casos como el desarrollo de sistemas de detección del fraude, como los sesgos en el resultado de salida de un tratamiento pueden ser necesarios para la medicina de precisión.

Las técnicas de “*active-learning*”, la representatividad, la relevancia y el poder predictivo no solo tienen ventajas para el desarrollo y el rendimiento de un modelo o sistema de IA, sino que también pueden ser una herramienta muy útil para aplicar el principio de minimización de datos.

E. CALIDAD DEL DATO INDIVIDUAL VS CALIDAD DEL CONJUNTO DE DATOS

Aunque se ha abordado de forma indirecta, conviene hacer explícita la distinción entre la calidad del dato, considerado individualmente, y la calidad del conjunto de datos¹³. Este último concepto adquiere especial relevancia en el caso del desarrollo de sistemas de IA basados en aprendizaje automático.

Por ejemplo, sea el caso de la siguiente tabla:

CONJUNTO DE DATOS				
ID	GENERO	PROVINCIA	SUELDO	EXP.
ID1	H	MAD	57400	Medio
ID2	H	MAD	67000	Bajo
ID3	H	MAD	90000	Alto
ID4	H	VAL	36900	Bajo
ID5	H	MAD	55000	Medio
ID6	H	MAD	76000	Alto
ID7	H	MAD	30000	Bajo
ID8	M	MAD	45000	Bajo

Figura 3 Conjunto de datos con posible cumplimiento el estándar ISO, interpretado a nivel de registro, pero claramente sesgado con relación al género.

La tabla anterior podría obtener un alto grado de calidad según el estándar ISO, interpretado a nivel de registro, pero el conjunto de datos está claramente sesgado con relación al género. Si la finalidad del tratamiento es tener estimaciones o tomar decisiones que afecten a todos los géneros, el conjunto de datos mostrado no sería adecuado para cumplir con la finalidad de forma idónea.

El tratamiento de datos personales que tiene como objetivo el desarrollo de sistemas de IA mediante aprendizaje automático es uno de entre muchos tratamientos en los que la consecución de la finalidad no depende tanto de la calidad del dato tomado individualmente, o del registro específico de un usuario, sino de la calidad del conjunto de datos.

ISO		Otras características	
Dato o registro	Conjunto de datos	Dato o registro	Conjunto de datos
Exactitud	Consistencia	Relevancia	Sesgo
Compleitud	Accesibilidad		Propiedades estadísticas
Credibilidad	Disponibilidad		Frecuencia
Actualidad	Portabilidad		Representatividad
Cumplimiento	Recuperabilidad		Entorno geográfico
Confidencialidad			Entorno contextual
Eficiencia			Entorno funcional
Precisión			Entorno conductual
Trazabilidad			Poder predictivo
Comprensibilidad			

Figura 4 Características específicas del dato o registro, y aquellas que dependen del conjunto de datos, y no las muestras evaluadas individualmente.

¹³ El empleo del término inglés “*data quality*”, puede ser interpretado tanto en singular como en plural, y puede llevar a equívocos.

Otros casos podrían ser tratamientos de análisis de tendencias, de correlaciones, optimizaciones, asignación de recursos, desarrollo de modelos, toma de decisiones en políticas públicas o análisis de riesgos en ámbitos como la logística, la educación, las finanzas, la energía, la movilidad, la sanidad, la ciencia, etc.

En estos casos, en la medida que el conjunto de datos contenga datos personales, el principio de exactitud-RGPD debe tener en cuenta características tan importantes como que los datos, tomados en su conjunto, no contengan sesgos que comprometan la finalidad del tratamiento.

Así pues, en sistemas de inteligencia artificial, la calidad del dato debe evaluarse también al nivel del conjunto de datos (“dataset”), pues resulta esencial para garantizar el correcto funcionamiento del sistema y, por lo tanto, cumplir con la finalidad del tratamiento que supone su desarrollo. Ello sin perjuicio de la calidad del dato individual, por cuanto produzca efectos sobre las personas.

El responsable del tratamiento, bajo el principio de exactitud y de responsabilidad proactiva, deberá garantizar y documentar la trazabilidad del conjunto de datos utilizado, incluyendo el origen de los datos, las transformaciones aplicadas y los criterios de selección y exclusión.

F. CALIDAD DE LOS DATOS VS VERACIDAD Y ACTUALIDAD

La calidad del conjunto de datos para una finalidad específica, en cuanto a datos personales, puede estar relacionada con la **veracidad y actualidad** de estos. No obstante, el RGPD es incluso explícito en que la actualidad del dato personal es un requisito de exactitud-RGPD sólo cuando sea necesario para cumplir con la finalidad del tratamiento.

La aplicación sobre un conjunto de datos personales de técnicas para la eliminación de sesgos, de técnicas de anonimización (no solo para anonimizar, sino para minimizar datos), de técnicas de privacidad diferencial o la generación de datos sintéticos puede implicar que algunos valores no reflejen directamente la realidad de un individuo o no cumplan determinados criterios técnicos de exactitud-ISO. Sin embargo, el conjunto de datos resultante puede tener la calidad suficiente como para ser utilizada como entrada en todas o algunas de las fases de desarrollo de un sistema de IA basado en ML. Es decir, el conjunto de datos puede considerarse que cumple con el principio de exactitud-RGPD en la medida en que resulte adecuado para la finalidad del tratamiento, sin que ello suponga prescindir de las exigencias de veracidad cuando estas sean relevantes.

Es más, existen técnicas para incrementar la calidad del conjunto de datos que pueden implicar la utilización de valores que no reflejen directamente la realidad individual. Así, por ejemplo, puede considerarse un conjunto de datos formado por registros con múltiples campos en el que algunos de ellos están vacíos o incompletos (problemas de completitud). Una técnica para aumentar la calidad del conjunto de datos es estimar el posible valor de los campos vacíos extrapolándolo de medidas estadísticas obtenidas del resto de registros. Esta es una técnica adecuada, pero introduce valores formalmente falsos e inexactos, sacrificando la exactitud-ISO por la completitud-ISO.

CONJUNTO DE DATOS ORIGINAL				COMPLETADO CON LA MEDIA			COMPLETADO CON MEDIA POR PROVINCIA		
ID	PROVINCIA	SUELDO	EXP.	PROVINCIA	SUELDO	EXP.	PROVINCIA	SUELDO	EXP.
ID1	MAD	57400	Medio	MAD	57400	Medio	MAD	57400	Medio
ID2	VAL	-	Bajo	VAL	65150	Bajo	VAL	22250	Bajo
ID3	GR	90000	Alto	GR	90000	Alto	GR	90000	Alto
ID4	VAL	36900	Bajo	VAL	36900	Bajo	VAL	36900	Bajo
ID5	BCN	-	Medio	BCN	65150	Medio	BCN	45000	Medio
ID6	VAL	7600	Alto	VAL	7600	Alto	VAL	7600	Alto
ID7	VAL	-	Bajo	VAL	65150	Bajo	VAL	22250	Bajo
ID8	BCN	45000	Bajo	BCN	45000	Bajo	BCN	45000	Bajo

Figura 5 En la tabla de la izquierda existen campos que no estaban completados en origen. Se puede completar la tabla de datos con distintas estrategias. En la tabla del centro se ha elegido introducir la media de los sueldos de toda la tabla. En la derecha se han utilizado las medias de los sueldos de la misma provincia.

Los conjuntos de datos personales que tienen campos “inventados” pueden mantener un nivel adecuado de calidad desde una perspectiva técnica. La característica de exactitud-ISO no se estaría cumpliendo para determinados registros individuales, pero eso no sería relevante para un tratamiento específico. Además, no se faltaría al principio de exactitud-RGPD si el conjunto de datos es adecuado para conseguir la finalidad del tratamiento.

De igual forma, una técnica de corrección de sesgos consiste en aplicar correcciones para normalizar por grupo, calculando el sesgo grupal y eliminando la diferencia sistemática:

TABLA ORIGINAL					TABLA CON SESGO ELIMINADO			
ID	Edad	Ingresos (€)	Código Postal	Historial Crédito	Edad	Ingresos (€)	Código Postal	Historial Crédito
1	45	60000	28001	0,80	45	60000	28001	0,80
2	25	28000	28003	0,70	25	28000	28003	0,90
3	50	72000	28002	0,85	50	72000	28002	0,85
4	38	58000	28001	0,90	38	58000	28001	0,90
5	29	30000	28003	0,60	29	30000	28003	0,80
6	41	65000	28002	0,85	41	65000	28002	0,85
7	33	32000	28003	0,65	33	32000	28003	0,85
8	41	65000	28002	0,95	41	65000	28002	0,95
9	36	70000	28001	0,75	36	70000	28001	0,75

Figura 6 En la tabla original hay un sesgo en la evaluación del historial de crédito con relación a las personas que viven en el código postal 28003. La media de los otros códigos postales es de 0,85, mientras que la media en las personas con código 28003 es de 0,65. El sesgo toma un valor de 0,2, que se puede corregir normalizando los valores.

Así pues, la calidad entendida como idoneidad para una finalidad, no se identifica automáticamente con la veracidad en sentido absoluto. La relevancia de la veracidad del dato varía en función del tipo de tratamiento. Así, en tratamientos agregados o estadísticos, pueden admitirse mayores niveles de abstracción o transformación del dato, mientras que en tratamientos que produzcan efectos sobre personas concretas, la exigencia de veracidad y actualidad, al menos en el resultado del mismo, debe ser más estricta. Ello sin perjuicio de que se debe garantizar que no se genere una representación incorrecta o distorsionada de la persona interesada, o que se generen efectos adversos sobre sus derechos e intereses. En definitiva, el uso de datos que no reflejen la realidad debe analizarse en función de su impacto en la toma de decisiones, la generación de perfiles, o cualquier otro efecto significativo sobre la persona física, y ha de estar justificado y documentado.

G. ALINEACIÓN DE EXACTITUD-RGPD CON CALIDAD DE LOS DATOS

El concepto de exactitud-RGPD se alinea con las definiciones de calidad de los datos de los estándares de la economía del dato en la medida en que ambos **están vinculados con el principio de finalidad**. Sin embargo, ambos términos no son intercambiables. Calidad de los datos se aplica tanto a datos personales como no personales en el marco de la industria, mientras que exactitud-RGPD se aplica a datos personales en el ámbito de competencia del RGPD¹⁴.

Por otro lado, el nivel de calidad de los datos puede venir determinado por requisitos técnicos o funcionales que no se derivan de la normativa de protección de datos, por ejemplo, el cumplimiento de obligaciones de otras normativas, cumplimiento de estándares, compromisos contractuales y políticas internas de una entidad. En consecuencia, dichos requisitos no tienen por qué coincidir con las exigencias del principio exactitud-RGPD, que deben valorarse en función de la finalidad del tratamiento y de sus efectos sobre las personas.

¹⁴ Como se ha señalado anteriormente, la referencia al RGPD empleada en el texto incluye a la L.O. 7/2021.

En ambos casos, los términos calidad de los datos y exactitud-RGPD no se configuran en términos absolutos, sino en relación con la finalidad del tratamiento. Por ello, **unos datos que resultan adecuados o exactos, y que por tanto que cumplen con la calidad para una finalidad, pueden no tener la calidad necesaria para otra finalidad distinta**. De igual forma, **los datos no veraces o no actuales, en la medida que permitan cumplir con la finalidad de un tratamiento específico, serían lo suficientemente exactos-RGPD o tendrían la necesaria calidad para dicho tratamiento**.

III. TRATAMIENTOS DE DATOS PERSONALES E IA

Un proceso es un conjunto de actividades mutuamente relacionadas, o que interactúan, las cuales transforman elementos de entrada en resultados de salida¹⁵. Todo proceso tiene un propósito o finalidad, y su implementación será eficaz cuando sea capaz de cumplir dicho propósito. Asimismo, para asegurar que dichos resultados cumplen unos **requisitos mínimos de eficacia**, deberán obtenerse **evidencias** mediante la **definición de métricas** específicas para cada tipo de proceso que permitan verificar si se alcanzan esos niveles.

En la medida en que el proceso implica datos personales nos encontramos ante un tratamiento-RGPD, que para cumplir el principio de finalidad ha de poder alcanzar su propósito, es decir, ser **idóneo**¹⁶.

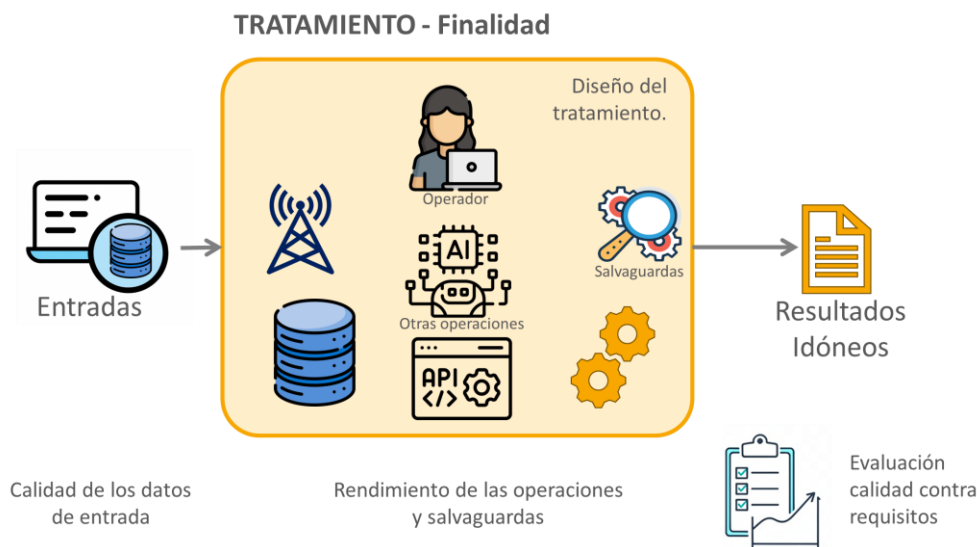


Figura 7 El resultado del tratamiento ha de ser idóneo respecto a su finalidad, lo que dependerá de su diseño, el rendimiento de las operaciones que lo forman, de las salvaguardas desplegadas y de los datos de entrada al tratamiento.

El cumplimiento de los requisitos de idoneidad dependerá, fundamentalmente, de los siguientes factores:

- El diseño del tratamiento como un conjunto de operaciones de distinta naturaleza¹⁷ y la relación entre ellas.
- La calidad de los datos de entrada al tratamiento, inclusive la calidad de los datos no personales.
- La calidad de los datos que sean resultados de cada una de las operaciones que conforman el tratamiento, que generarán datos intermedios.

¹⁵ ISO 9001:2015

¹⁶ STC 14/2003, de 28 de enero, juicio de idoneidad: evaluar si la medida es susceptible de conseguir el objetivo propuesto.

¹⁷ Operaciones ejecutadas por personas, por sistemas digitales, por sistemas de IA, por el responsable, por un encargado, en la nube, automatizadas, etc.

- Las salvaguardas desplegadas en el tratamiento (p.ej. preprocesado de los datos, supervisión humana, supervisión con algoritmos, evaluación de convergencia, filtrado de acuerdo a políticas, alertas, gestión de quejas e incidencias, ...), que podrían aplicarse para corregir problemas de calidad de los datos de entrada, problemas de calidad en los datos intermedios generados por cada uno de los sistemas que implementan las operaciones de tratamiento y, también, las salvaguardas que corrigen el resultado de salida del tratamiento.

Estos factores son interdependientes, y han de estar adecuadamente balanceados para conseguir el grado de idoneidad deseado en el resultado del tratamiento.

A. TIPOS DE TRATAMIENTOS DE DATOS PERSONALES

El resultado del tratamiento de datos, tanto personales como no personales, puede ser una acción, una decisión, incluyendo una decisión sobre una persona física, un perfil sobre dicha persona o un dato personal en general. También puede dar como resultado (incluso simultáneamente) datos no personales, como una decisión estratégica, un valor estadístico, la definición de una política o un modelo de IA que no contenga datos personales.

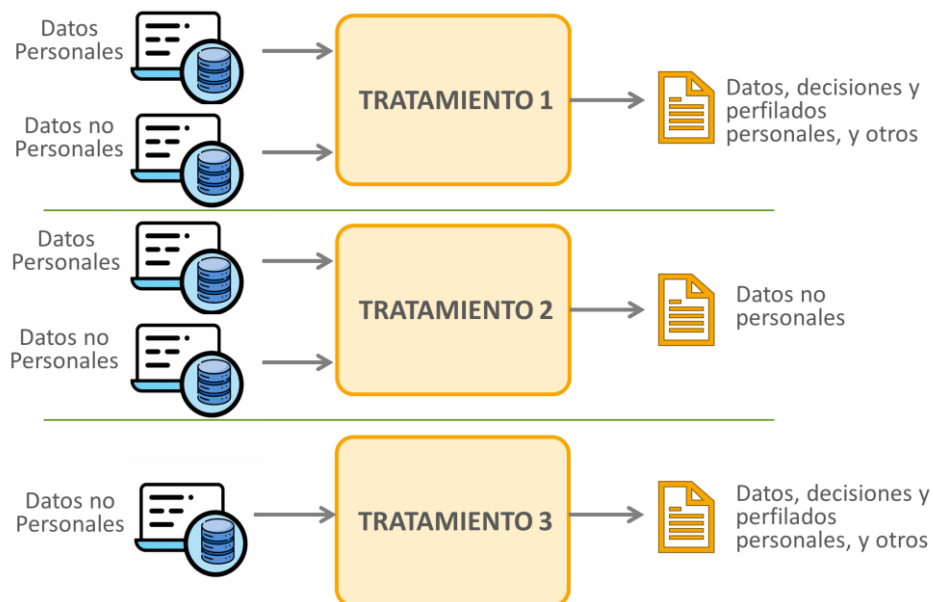


Figura 8 Tres casos de tratamientos de datos personales clasificados con el criterio de si los datos de entrada al tratamiento son personales o no, y si el resultado de salida es un dato personal o no.

Un tratamiento (caso 1) podrá, en función de datos personales y no personales (estos últimos podría ser, por ejemplo, baremos, estadísticas, presupuestos u otros) obtener un resultado de salida que podría determinar si un ciudadano va a recibir una beca o una ayuda, evaluar su fiabilidad financiera, calificarlo para un trabajo, determinar su productividad, darle acceso a un servicio o local, etc. Es decir, tomar una decisión, perfilarlo o inferir otros datos. Este resultado de salida, que suponen datos vinculados a una persona física, ha de ser lo suficientemente exactos-RGPD (y mínimos-RGPD) para que la finalidad del tratamiento se cumpla de forma idónea. A su vez, los datos personales de entrada al tratamiento tendrán que ser exactos-RGPD (y mínimos-RGPD). También, el conjunto de todos los datos de entrada al tratamiento, incluyendo los no personales, tendrá que tener la calidad de los datos suficiente para que la finalidad se pueda alcanzar tal y como está diseñado e implementado el tratamiento.

Un tratamiento (caso 2) puede utilizar datos personales (y no personales) para definir una política pública, como la distribución de recursos educativos en función del perfil y rendimiento de los alumnos, analizar una situación económica de una región, etc. En este caso el resultado de salida de dicho tratamiento no sería un dato personal. Sin embargo, el resultado de salida deberá

tener la calidad suficiente para que la finalidad se pueda cumplir y el tratamiento de los datos personales utilizados sea leal y legítimo. A su vez, los datos de entrada al tratamiento tendrían que cumplir las condiciones descritas en el caso anterior.

Un tratamiento (caso 3) que sin recibir de entrada ningún dato personal tome una decisión sobre una persona física, perfile o infiera datos que se vinculen a una persona física es más difícil de plantear. Podría ser el caso de un sorteo de plazas, la generación de números en una lotería, o cualquier otro caso en el que se aplique una regla ajena al individuo pero que finalmente se asocie a él o le afecte significativamente. En la medida que el resultado sean datos personales tendrá que cumplir las mismas condiciones que en los casos anteriores.

B. CALIDAD DE LOS DATOS EN TRATAMIENTOS DE DATOS PERSONALES QUE INCLUYEN SISTEMAS DE IA

Las distintas operaciones de un tratamiento concreto se pueden implementar de diferentes formas: con personas y/o con sistemas y/o, en particular, con uno o más sistemas de IA. Si nos enfocamos en el sistema de IA, este ha de producir resultados con la calidad suficiente para cumplir la finalidad del tratamiento.

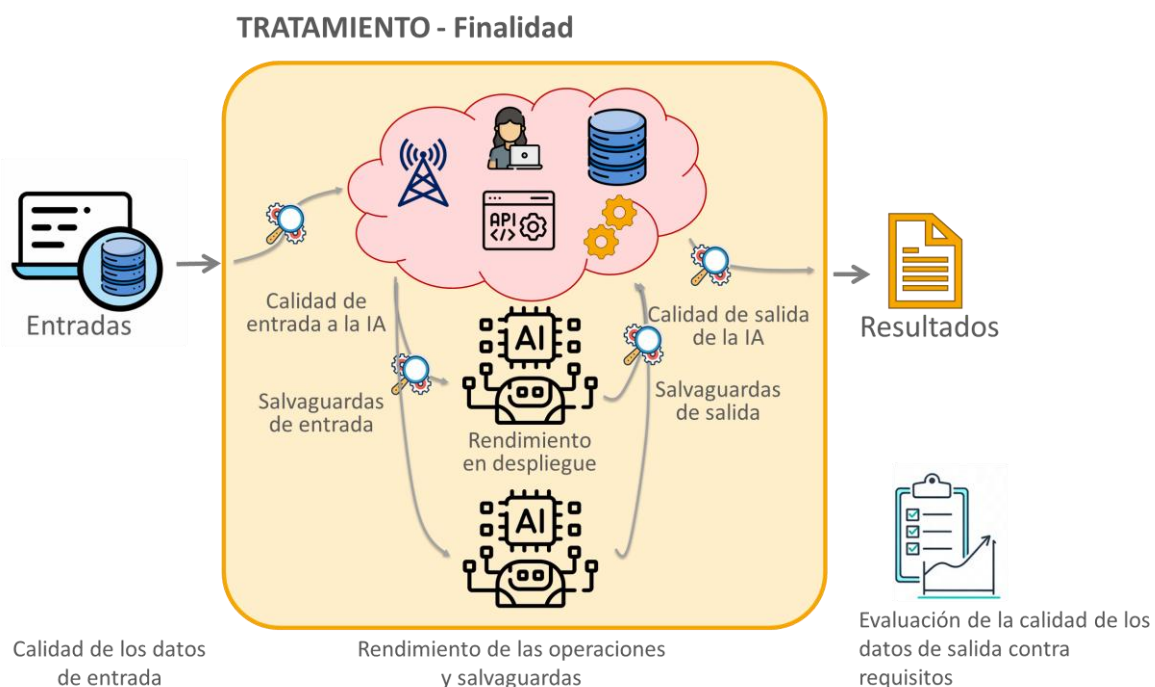


Figura 9 Aspectos de calidad de datos para tener en cuenta cuando se despliega uno o más sistemas/modelos de IA en un tratamiento. En la figura se han representado deliberadamente dos, para evidenciar que no hay una correspondencia directa entre sistema de IA y tratamiento

El sistema de IA podría ser tan simple como un modelo de IA con los mínimos interfaces o, por el contrario, un complejo servicio SaaS de IA generativa, o agéntica, que ya incorpora numerosas operaciones y salvaguardas¹⁸ tanto a la entrada como a la salida del sistema.

Los resultados de salida de un sistema de IA (ya se denominen inferencias¹⁹, predicciones, contenidos, recomendaciones o decisiones, entre otros) que forma parte de un tratamiento podrán ser una entrada para operaciones posteriores del mismo tratamiento. Entre ellas, y sería

¹⁸ Firewalls, limitadores de interacción, balanceo de carga, circuit breakers, detectores de información personal, clasificadores de texto y contenido, higienizadores, modelos de embedding, rerankers, memorias, RAGs y su gestión, generador del prompt, selectores de modelos, escudo, autorecuperación, reformato, detector de alucinaciones, de copyright, gestor de referencias y citas, filtros de toxicidad, aplicación de políticas de servicio, llamadas a herramientas, APIs, clientes MCP, etc.

¹⁹ El RIA, en el considerando 12, habla de "capacidad de inferencia" como la capacidad de obtener los resultados de salida enumerados a continuación.

lo más recomendable, operaciones que implementen salvaguardas adicionales y supervisión humana. Por ello, el nivel mínimo de calidad de los datos y/o exactitud-RGPD exigible a los resultados de salida del sistema IA dependerá de varios factores:

- Los requisitos de calidad de datos y/o exactitud-RGPD definidos para garantizar el cumplimiento de la finalidad del tratamiento de forma idónea.
- El respeto a los derechos y libertades de las personas físicas identificadas o identificables cuando dichos resultados sean relativos a ellas.
- El conjunto de operaciones posteriores al sistema IA dentro del tratamiento que modifiquen o corrijan los resultados generados por dicho sistema de IA (un sistema IA podría generar salidas para otro sistema IA dentro del mismo tratamiento, como en los sistemas multiagente).
- En particular, de las salvaguardas internas incluidas en el diseño del tratamiento destinadas a corregir los errores conocidos del sistema de IA.

La capacidad del sistema de IA de proporcionar la calidad exigible una vez desplegado en el tratamiento no dependerá únicamente de éste, sino de varios factores de diseño del tratamiento y de su contexto, por ejemplo:

- De la calidad de los datos de entrada al sistema IA, que podrán haber sido resultados de otras operaciones que se realizan en el marco del tratamiento, como recogida de datos de sensores, bases de datos o interacción con un humano.
- De las salvaguardas desplegadas durante el diseño del tratamiento que se apliquen a dichos datos de entrada, como las que impiden una degradación en la dinámica de la operación sobre el sistema IA desplegado (bucle de retroalimentación²⁰).
- De la adecuación del sistema IA, diseñado para unos propósitos más o menos concretos, con relación a la finalidad del tratamiento donde se ha desplegado.

Los requisitos de calidad de los datos de entrada al tratamiento podrán definirse cuando se conozcan las condiciones de idoneidad del tratamiento. Estas, a su vez, se podrán determinar teniendo en cuenta, entre otros, los requisitos de exactitud-RGPD del resultado y el impacto a los derechos de las personas físicas. El diseño del tratamiento, incluyendo las salvaguardas que se planifiquen, podrán matizar dichos requisitos de calidad de los datos de entrada.

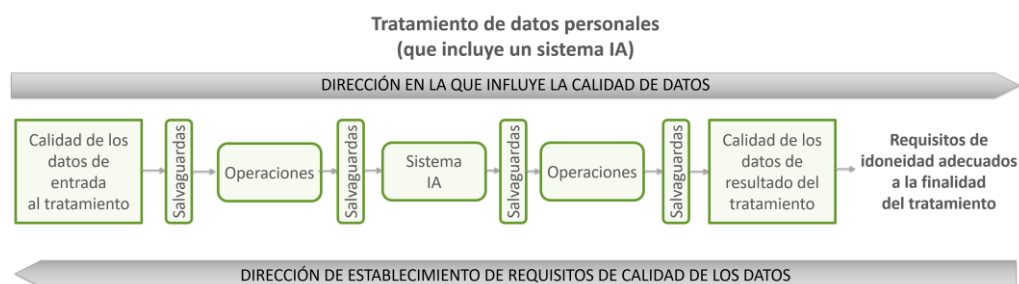


Figura 10 Definir los requisitos de calidad de los datos de entrada no es posible si no se conocen los requisitos de idoneidad de los tratamientos probables.

Los hechos demuestran que el sistema perfecto es una ilusión, ya sea sistema IA o no-IA. La adecuación de un sistema IA dependerá de cómo el diseño del tratamiento, donde se despliega, aproveche sus bondades y gestione adecuadamente sus defectos en el marco de un contexto y una finalidad específica. Un sistema o modelo de IA puede tener muy distinto rendimiento dependiendo de las condiciones de despliegue en un entorno determinado. Por lo tanto, **no sería exigible un sistema IA perfecto, sino un sistema IA adecuado al diseño del tratamiento**, teniendo en cuenta los efectos del sistema sobre las decisiones o evaluaciones sobre las personas

²⁰ Considerando 67 del RIA. En el caso de despliegue sin reentrenamiento, porque se alimenten RAGs, prompts o memoria de forma dinámica.

afectadas. Debe destacarse que el rendimiento agregado del sistema de IA puede ocultar errores o inexactitudes en inferencias individuales, especialmente cuando se evalúa mediante métricas globales. Ello exige la adopción de mecanismos adicionales de control cuando los resultados del sistema puedan afectar a personas concretas.

C. EXIGIBILIDAD DE CALIDAD EN LOS DATOS NO PERSONALES DENTRO DE TRATAMIENTOS DE DATOS PERSONALES

En muchos casos, en un tratamiento de datos personales también estarán involucrados datos no personales, tanto entradas al tratamiento como formando parte del diseño del mismo²¹. La calidad de los datos no personales también puede ser crítica para conseguir la idoneidad del tratamiento y/o con relación al impacto a los derechos y libertades de personas físicas identificadas o identificables. En esa medida, se deben establecer los requisitos de calidad de los datos no personales y medidas adecuadas para garantizar el cumplimiento del tratamiento.

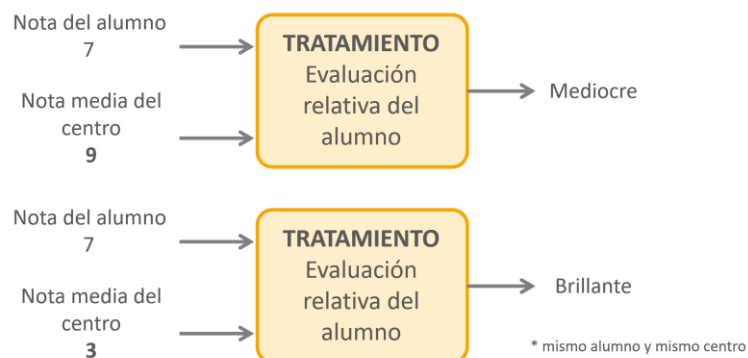


Figura 11 Ejemplo de cómo un dato no personal de entrada puede influir en el resultado del tratamiento y afectar a una persona. Un error (una falta de calidad) en el dato no personal de entrada puede generar un resultado de salida no exacto-RGPD, aunque el dato personal de entrada sea exacto-RGPD.

D. DESARROLLO DE IA COMO TRATAMIENTO

Un caso específico de tratamiento de datos personales (ver caso 2 definido en el apartado A de este capítulo) involucra el desarrollo o evolución de un modelo o sistema de IA con aprendizaje automático basado en datos. Dicho desarrollo será un tratamiento distinto del tratamiento en el que el sistema de IA es utilizado en fase de explotación o despliegue para tratar datos personales.

La finalidad del tratamiento de desarrollo podría ser crear un sistema que realice una tarea concreta, y ya en el marco del tratamiento decidir que la naturaleza de dicho sistema será una IA. O bien la finalidad podría ser el desarrollo de una IA en sí misma, por ejemplo, en el caso de proyectos cuya financiación o subvención requiera el diseño específicamente de un modelo de IA.

El ciclo de vida de dicho tratamiento se podría representar utilizando las fases de desarrollo y evolución definidas en el estándar ISO 5338.

²¹ Pueden tomar la forma de bases de datos, ficheros, tablas y ciertos parámetros de configuración de los sistemas empleados en la implementación del tratamiento.

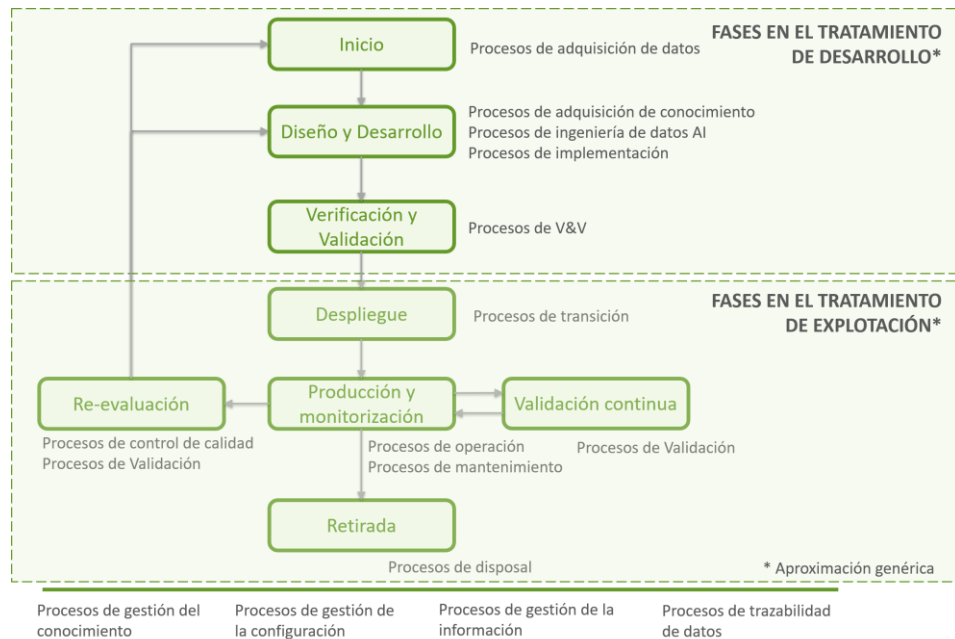


Figura 12 Fases del ciclo de vida de desarrollo de sistemas IA acorde al estándar ISO 5338 destacando los procesos de datos y una clasificación con carácter general de las fases del ciclo de vida que formarían parte de operaciones de un tratamiento de desarrollo y de aquella que formarían parte de un tratamiento de explotación cuando suponga tratamiento de datos personales.

El rendimiento del sistema o modelo IA resultante, con relación a la calidad de los resultados que ofrezca, dependerá, entre otros, de:

- La calidad de los **conjuntos de datos** de desarrollo (entrenamiento, prueba, validación, evolución).
- La modificación de la calidad de los datos a lo largo de los procesos de datos del ciclo de vida.
- La metodología de desarrollo empleada.
- Las decisiones de diseño.
- Los contextos y finalidades previstas.
- Los factores de implementación en el despliegue (p.ej. cuantización del modelo).
- La evolución del contexto en el tiempo (p.ej. casos de context-drift y data-drift)
- En el caso de evolución del modelo de IA, por ejemplo, sesgos generados por bucles de realimentación.

E. DETERMINACIÓN DE LOS REQUISITOS DE CALIDAD DE LOS DATOS EN EL DESARROLLO DE SISTEMAS DE IA

La gestión de la calidad de los datos aplicable a un tratamiento destinado al desarrollo de un sistema de IA, tal y como se ha señalado en el apartado B de este capítulo, solo podrá definirse una vez que se conozcan las condiciones de idoneidad de los procesos o tratamientos en los que se prevé desplegar el sistema, puesto que dichas condiciones determinan los requisitos de calidad exigibles a los datos.

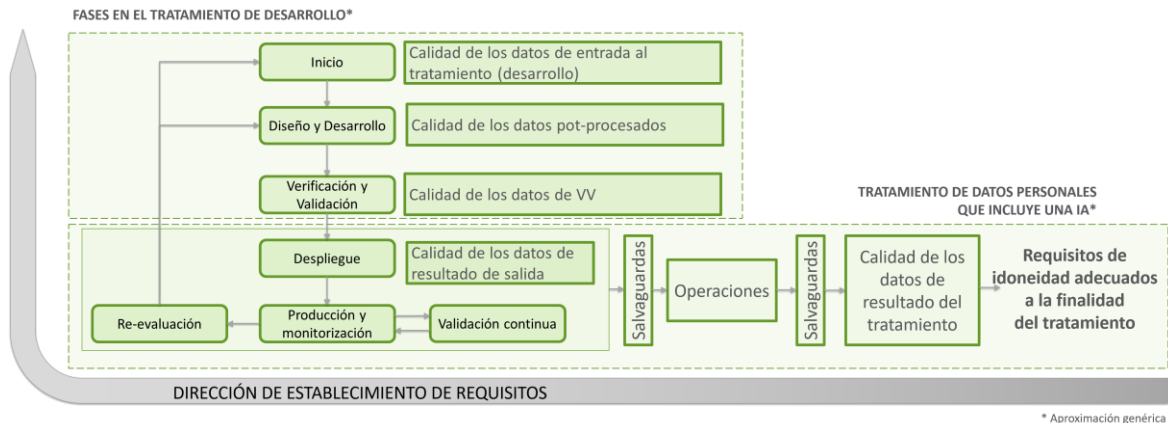


Figura 13 La calidad de los datos de los resultados de la IA será un requisito que se derivará de la calidad necesaria de los resultados en los tratamientos previstos.

También podría ser necesario conocer en avance, o establecer como un requisito de despliegue del sistema IA, las salvaguardas que deberán incluirse en el tratamiento y el rendimiento de otros sistemas, IA o no-IA, que interaccionarán con el primero.

La calidad de los datos no es un requisito estático fijado en la fase de diseño. Hay que entenderlo como un proceso de gestión continuo a lo largo del ciclo de vida del modelo o el sistema de IA y, también, del ciclo de vida de los tratamientos de datos personales. Para ello se requiere la monitorización, revisión y, en su caso, actualización de los conjuntos de datos y de los sistemas que los utilizan, especialmente cuando puedan producirse cambios en el contexto, en los datos de entrada o en el comportamiento del sistema (gobierno del dato).

F. ANÁLISIS DE CALIDAD DE LAS FUENTES DE DATOS PARA IA

Para el desarrollo de sistemas de IA es necesario el acceso a conjuntos de datos con la calidad suficiente. El acceso a datos, como lo define la DGA en su art.2.13 es “toda **utilización** de datos de conformidad con unos **requisitos específicos** de carácter **técnico, jurídico u organizativo, sin que ello implique necesariamente la transmisión o la descarga de los datos;**”.

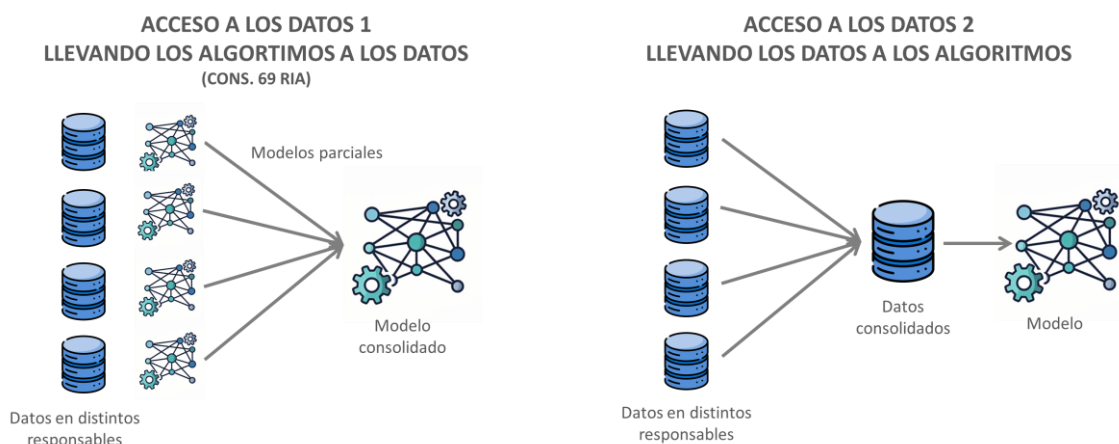


Figura 14 A la izquierda se encuentra el modelo de acceso basado en llevar los algoritmos a los datos “*compute to data*”, que en desarrollo de sistemas de IA se denomina aprendizaje federado. Sus ventajas no se limitan a la protección de datos, sino en garantizar la soberanía del dato. En algunos casos, este modelo de acceso es obligado, por ejemplo, en el caso los usos secundarios en el Espacio Europeo de Datos de Salud (entornos de tratamiento seguro).

El acceso a datos puede constituir un tratamiento autónomo —por ejemplo, para la creación de catálogos de datos— o integrarse como una operación dentro de un tratamiento destinado al desarrollo de un sistema de IA.

Antes de iniciar un proceso de acceso o recogida masiva de datos, deberá evaluarse la legitimación de dicho acceso. Para ello se deberá determinar, entre otros, si los conjuntos de datos son **necesarios**, es decir, si son **adecuados, pertinentes y limitados** a lo **necesario** en relación con los **finés**, como establece el principio de minimización-RGPD. Esto implica que habría que hacer una evaluación de la calidad de los datos en origen y evaluar si tienen la calidad suficiente, por sí mismos o en conjunción con otros conjuntos de datos.

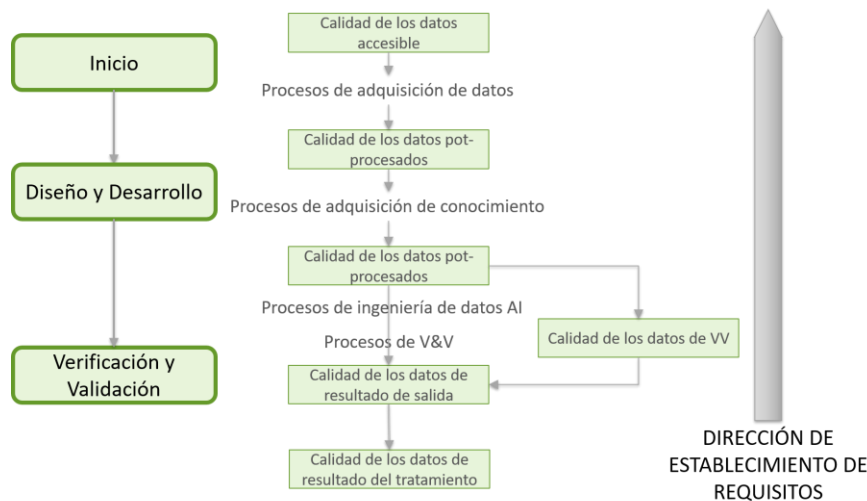


Figura 15 La determinación de los requisitos de rendimiento del sistema IA en los contextos de uso previstos es el punto de inicio para evaluar cuál es la necesaria calidad de los datos a los que tener acceso.

Este análisis cabría hacerlo tanto en el marco del acceso a datos para desarrollo de sistemas o modelos de IA, como para la ejecución de cualquier otro tratamiento.

G. REQUISITOS DE CALIDAD INADECUADOS

La **definición de los requisitos de calidad** de los datos sobre la base de **criterios objetivos y evidencia verificable** constituye un elemento esencial para el desarrollo de sistemas y modelos de IA. En efecto, el establecimiento de unos **requisitos de calidad insuficientes** podría **comprometer la idoneidad** del tratamiento de desarrollo, al impedir que el sistema finalmente obtenido alcance las prestaciones o satisfaga los requisitos de diseño previstos.

Por otro lado, identificar unos **requisitos de calidad demasiado altos** no solo **contradice el principio de minimización-RGPD** al implicar el **acceso a un volumen mayor de datos** o a datos con mayor extensión, granularidad, precisión o frecuencia de lo necesitado. Además, **encarecería el acceso** a los conjuntos de datos, a **la gestión** de los procesos de gobierno del dato²², y **otros procesos de desarrollo** de los sistemas de IA. En el peor de los casos, ello **podría llevar a descartar iniciativas** empresariales o de investigación por la errónea asunción de que no se dispone de un conjunto de datos con la calidad suficiente o de que no es posible obtenerlo (por otras razones legales, económicas o de intereses comerciales).

IV. REFLEXIONES FINALES

Como conclusión, se formulan las siguientes reflexiones sobre la calidad de los datos y el principio de exactitud-RGPD en los tratamientos, tanto en aquellos en los que se despliegan sistemas y modelos de IA como en los destinados a su desarrollo o evolución:

- Cuando el resultado de un tratamiento consista en una decisión, un perfilado, una inferencia o un enriquecimiento de información relativo a una persona física identificada o

²² Por ejemplo, de los 22 procesos que, para el gobierno, la gestión y la gestión de la calidad del dato, se desarrollan en la norma UNE 0080:2023.

identificable, dicho resultado tendrá la consideración de dato personal. En consecuencia, deberá cumplir los principios de exactitud y minimización previstos en el RGPD, en relación con la finalidad del tratamiento, garantizando en todo caso el respeto de los derechos y libertades de las personas físicas²³.

- Los datos personales utilizados como entrada en un tratamiento deberán cumplir igualmente los principios de exactitud y minimización previstos en el RGPD, en la medida en que ello resulte necesario para garantizar el cumplimiento de la finalidad del tratamiento y el respeto de los derechos y libertades de las personas físicas cuando el resultado de dicho tratamiento les afecte.
- El principio de exactitud previsto en el RGPD exige que los datos personales utilizados como entrada en un tratamiento reúnan el grado de calidad necesario para garantizar su idoneidad respecto de la finalidad perseguida. Ello puede requerir distintos niveles de veracidad, actualización, precisión, granularidad, frecuencia u otros atributos. En consecuencia, el principio de exactitud no se limita a exigir la veracidad y la actualización de los datos, sino que estas deberán garantizarse únicamente en la medida en que resulten necesarias para cumplir la finalidad del tratamiento, especialmente puedan incidir en las personas físicas, pero no en todos los casos.
- En aplicación del principio de minimización previsto en el RGPD, únicamente podrán tratarse los datos personales de entrada que sean adecuados, pertinentes y limitados a lo necesario para la finalidad del tratamiento. En consecuencia, el RGPD no legitima el tratamiento de datos personales con un grado de precisión, granularidad, frecuencia u otras características que excedan de lo necesario para el cumplimiento de dicha finalidad.
- Los principios de exactitud y minimización previstos en el RGPD delimitan el rango de calidad que deben reunir los datos personales para garantizar el cumplimiento de la finalidad del tratamiento y la adecuada realización de las distintas operaciones que lo integran.
- Los requisitos de calidad de los datos y el principio de exactitud previsto en el RGPD son conceptos estrechamente relacionados, pero no equivalentes. La calidad de los datos posee un alcance más amplio, ya que se extiende también a los datos no personales y comprende requisitos que trascienden el cumplimiento de la normativa de protección de datos.
- En el marco de un tratamiento de datos personales, los datos no personales utilizados deberán cumplir asimismo los requisitos de calidad que resulten adecuados, en la medida en que su utilización pueda afectar a la idoneidad del tratamiento con relación a sus fines, o a los resultados del tratamiento cuando estos puedan incidir en los derechos y libertades de las personas físicas.
- Debe asumirse que no existen tratamientos perfectos. Deficiencias en la recogida de los datos de entrada, errores en las operaciones de tratamiento, deficiencias de diseño u otras incidencias pueden degradar la calidad e idoneidad de los resultados obtenidos. Por ello, es necesario determinar el grado de calidad mínimo aceptable que deben reunir los resultados del tratamiento para cumplir su finalidad, definir métricas que permitan evaluarlo e incorporar al diseño del tratamiento mecanismos de seguimiento, control y gestión de incidencias que garanticen la idoneidad de dicho tratamiento.
- Tampoco existe un sistema perfecto para implementar las operaciones de tratamiento, en particular cuando se emplean sistemas de IA. No obstante, sí existe un sistema de IA adecuado para su despliegue en un tratamiento cuando sus fortalezas y debilidades son conocidas y están debidamente documentadas, y estas últimas pueden gestionarse eficazmente mediante el diseño e implantación de las salvaguardas necesarias en dicho tratamiento.

²³ Aunque ya se ha señalado anteriormente, en particular, cuando el resultado sea relativo a una persona identificada o identificable y en relación, entre otros, del cumplimiento de los principios de exactitud y minimización.

- En un tratamiento, el ignorar el uso de métricas de calidad del resultado de salida impide garantizar y poder demostrar²⁴ si realmente se están cumpliendo con eficacia e idoneidad los fines del tratamiento.
- La determinación del nivel mínimo de calidad de los datos necesario para que el resultado del tratamiento sea idóneo permite adoptar decisiones racionales sobre el diseño del tratamiento, sus distintas operaciones, la relación entre ellas, la selección de los sistemas más adecuados para su ejecución en función de su rendimiento y las salvaguardas que deban implantarse. En particular, en relación con los sistemas de IA.
- En los tratamientos de aprendizaje automático que utilicen datos personales, el conjunto de datos empleado deberá reunir la calidad y la exactitud conforme al RGPD que posibilite el desarrollo de un sistema de IA que cumpla los requisitos de rendimiento previstos para su despliegue.
- En el aprendizaje automático, la calidad del conjunto de datos, considerada de forma agregada, resulta más relevante que las características de calidad de cada dato individualmente considerado.
- Para determinar la calidad necesaria del conjunto de datos para el desarrollo o la evolución de un sistema de IA, deben tenerse en cuenta, entre otros factores, el preprocesamiento de los datos, la metodología de desarrollo, el tipo de despliegue, el contexto de uso y las salvaguardas previstas para su integración en el tratamiento.
- La calidad de los datos de salida de un sistema IA, necesaria para que cumpla los requisitos de rendimiento cuando se despliega en un tratamiento, es distinta de la calidad exigible a los datos utilizados durante su desarrollo. En consecuencia, ambas deben evaluarse de forma diferenciada.
- Los conjuntos de datos que no reúnan la calidad necesaria para el desarrollo o la evolución de un sistema o modelo de IA no serán necesarios para dicho tratamiento. En consecuencia, no estará justificado el acceso a ellos, salvo en la medida estrictamente necesaria para evaluar su calidad.
- El establecimiento de requisitos de calidad de los datos excesivos para tratamientos destinados al desarrollo de sistemas de IA puede comprometer el cumplimiento del principio de minimización establecido en el RGPD, incrementar innecesariamente los costes del tratamiento de desarrollo y, en el peor de los casos, provocar la cancelación de iniciativas o proyectos de investigación.
- La gobernanza de datos de la organización, así como los procesos de gestión que de ella se deriven, deberán integrar los requisitos de gobernanza aplicables tanto a los tratamientos de datos personales como a los procesos de desarrollo, adquisición y evolución de sistemas de IA.
- Los procesos de gobernanza de datos deberán definir de forma clara quién y cómo se determinarán de los requisitos de calidad de los datos en la organización responsable. Dichos requisitos deberán definirse de forma objetiva y garantizar el cumplimiento de los principios de exactitud y minimización establecidos en el RGPD.
- La calidad de los datos deberá ser objeto de seguimiento continuo a lo largo del ciclo de vida del tratamiento, mediante la incorporación de mecanismos de revisión periódica y de adaptación a los cambios en el contexto, las finalidades, el alcance y los riesgos del tratamiento, así como a las variaciones en el comportamiento de los sistemas utilizados.
- En el diseño de un tratamiento, la documentación de los requisitos relativos a la calidad de los datos, así como del cumplimiento de los principios de exactitud y minimización establecidos en el RGPD, constituye un elemento esencial del principio de responsabilidad

²⁴ Artículo 24 del RGPD

proactiva. En el caso de los tratamientos destinados al desarrollo de sistemas de IA, dicha documentación deberá tener en cuenta los usos previsibles de los sistemas.

- El responsable del tratamiento, al seleccionar los sistemas de IA que vayan a utilizarse para implementar operaciones de tratamiento, tendrá en cuenta los requisitos documentados por el proveedor en relación con la calidad de los datos y, en su caso, las salvaguardas y demás aspectos que deban incorporarse al diseño del tratamiento para garantizar un despliegue adecuado.
- Un DPD o un asesor en protección de datos debe conocer y comprender las implicaciones de todos los flujos y operaciones sobre datos que se producen en tratamientos de desarrollo de sistemas y modelos de IA, así como los que se producen en tratamientos que incorporan IA. Este conocimiento permitirá al DPD y al asesor proporcionar soluciones y contribuir a resolver los dilemas que puedan obstaculizar tanto la innovación como la protección de los derechos fundamentales.
- En el desarrollo de sistemas y modelos de IA basados en datos, es fundamental disponer de un equipo que incorpore, al menos, científicos de datos y especialistas en protección de datos. Este equipo debe establecer requisitos de calidad de los datos basados en criterios objetivos y evidencia verificable a lo largo de todo el ciclo de vida de los tratamientos de datos personales. De este modo, se contribuye a garantizar tanto el éxito de los proyectos como la protección de los derechos y libertades fundamentales.

Finalmente, cabe señalar que esta nota técnica está relacionada con otro material publicado por la AEPD en materia de inteligencia artificial y protección de datos, entre otros:

- [Inteligencia Artificial Agéntica desde la perspectiva de Protección de Datos](#) [feb 2026]
- [Requisitos para Auditorías de Tratamientos que incluyan IA](#) [ene 2021]
- [Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial](#) [feb 2020]

V. ANEXO: EVOLUCIÓN DEL CONCEPTO DE EXACTITUD EN LA NORMATIVA ESPAÑOLA

La Directiva 95/46/CE, en cuanto a los datos personales, establecía en su sección primera, “*Principios relativos a la calidad de los datos*” el concepto de calidad de datos que los datos debían ser:

*c) adecuados, pertinentes y no excesivos **con relación a los fines** para los que se recaben y para los que se traten posteriormente;*

*d) **exactos** y, **cuando sea necesario, actualizados**; deberán tomarse todas las medidas razonables para que los datos inexactos o incompletos, **con respecto a los fines** para los que fueron recogidos o para los que fueron tratados posteriormente, sean suprimidos o rectificados;*

La Directiva vinculaba los conceptos de exactitud y minimización para ser interpretados en el marco de la finalidad del tratamiento. La Ley Orgánica 15/1999, que trasponía la Directiva, en el TÍTULO II “Principios de la protección de datos”, en su art.4 “*Calidad de los datos*”, enunciaban una serie de principios, cercanos a los establecidos en el art.5.1 del RGPD. En particular:

*1. Los datos de carácter personal **sólo** se podrán recoger **para su tratamiento**, así como someterlos a dicho tratamiento, **cuando sean adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades** determinadas, explícitas y legítimas para las que se hayan obtenido*

*3. Los datos de carácter personal serán **exactos** y puestos al día de forma que **respondan con veracidad a la situación actual del afectado***

La redacción de la LO 15/1999 desvincula el concepto de exactitud de datos al de la necesidad para cumplir finalidad del tratamiento, definiendo la exactitud del dato personal de forma absoluta con relación a la veracidad de la situación actual del afectado. De dicha redacción se podría concluir:

- Que el dato se podría considerar exacto, aunque afectase a la finalidad del tratamiento.
- Y que para que el dato para ser considerado exacto debía responder con veracidad a la situación actual del afectado, aunque dicha circunstancia no afectase a la finalidad del tratamiento y, por tanto, no fuese necesaria.

El RGPD no hace referencia al término “*calidad de los datos*”²⁵, aunque si se establece los principios de minimización y de exactitud en sus art.5.1.c) y d):

“1. Los datos personales serán:

*c) **adecuados, pertinentes y limitados** a lo necesario **en relación con los fines** para los que son tratados («minimización de datos»);*

*d) **exactos** y, si fuera **necesario, actualizados**; se adoptarán todas las medidas razonables para que se supriman o rectifiquen sin dilación los datos personales que sean inexactos **con respecto a los fines para los que se tratan** («exactitud»);”.*

El RGPD se alinea con las definiciones de la Directiva vinculando minimización y exactitud a la finalidad del tratamiento.

En la LOPDGDD, en su art.4, la redacción es distinta, aunque haciendo referencia explícita al artículo 5.1.d):

Artículo 4. Exactitud de los datos

*1. Conforme al artículo 5.1.d) del Reglamento (UE) 2016/679 los datos serán **exactos** y, si fuere necesario, actualizados.*

*2. A los efectos previstos en el artículo 5.1 d) del Reglamento (UE) 2016/679, no será imputable al responsable del tratamiento, siempre que éste haya adoptado todas las medidas razonables para que se supriman o rectifiquen sin dilación, **la inexactitud de los datos personales, con respecto a los fines para los que se tratan**, cuando los datos inexactos:*

a) Hubiesen sido obtenidos por el responsable directamente del afectado.

b) Hubiesen sido obtenidos por el responsable de un mediador o intermediario en caso de que las normas aplicables al sector de actividad al que pertenezca el responsable del tratamiento establecieran la posibilidad de intervención de un intermediario o mediador que recoja en nombre propio los datos de los afectados para su transmisión al responsable. El mediador o intermediario asumirá las responsabilidades que pudieran derivarse en el supuesto de comunicación al responsable de datos que no se correspondan con los facilitados por el afectado.

c) Fuesen sometidos a tratamiento por el responsable por haberlos recibido de otro responsable en virtud del ejercicio por el afectado del derecho a la portabilidad conforme al artículo 20 del Reglamento (UE) 2016/679 y lo previsto en esta ley orgánica.

d) Fuesen obtenidos de un registro público por el responsable.

La LOPDGDD exige exactitud sin definir qué es exactitud, aunque la referencia al 5.1.d) debería entenderse que habría de estar vinculado al cumplimiento de los fines. Sin embargo, tal vez debido

²⁵ Excepto en el art.47 “Normas corporativas vinculantes” en el apartado 2 letra d) “Las normas corporativas vinculantes mencionadas en el apartado 1 especificarán, como mínimo, los siguientes elementos: d) la aplicación de los principios generales en materia de protección de datos, en particular la limitación de la finalidad, la minimización de los datos, los periodos de conservación limitados, la calidad de los datos, la protección de los datos desde el diseño y por defecto, ...”.

al sesgo de resistencia cognitiva²⁶, en muchos casos se arrastra el concepto definido en el marco de la L.O. 15/99 de que “exacto” significa responder con veracidad a la situación actual del afectado.

Dicha interpretación podría afectar:

- A los derechos de los interesados, en la medida que exige datos que podrían ir más allá de las necesidades del tratamiento, contradiciendo el principio de minimización.
- A los derechos de los interesados, ya que la redacción de la LOPDGDD permitiría que se empleen, en ciertos casos, datos que no son adecuados para la finalidad del tratamiento y, por tanto, contradicen lo establecido en el art.5.1.d) del RGPD.
- A los responsables y encargados, y en particular a los desarrolladores/evolucionadores de sistemas IA, al exigirles unos niveles de veracidad y actualidad que podrían no ser necesarios, podrían ir en contra de la propia calidad de los datos para desarrollar sistemas, y además de exigirles la atención de unos derechos más allá de lo establecido en el RGPD²⁷.

²⁶ O sesgo de perseverancia, la resistencia a aceptar un cambio de las ideas aunque haya evidencia para ello.

²⁷ El art.16 del RGPD también vincula la rectificación a la finalidad del tratamiento.