

REVISIÓN ACADÉMICA Y ESTRATÉGICA

Inteligencia artificial y valor estratégico de los datos en las industrias alimentaria, químico-farmacéutica y biotecnológica

De la acumulación informacional a la evidencia científica, el capital industrial y la ventaja competitiva defendible



José M. López, BSc,MD,PhD, LLD (Cand.)

Dirección Técnica e I+D+i · INTABIOTECH S.L.

Julio de 2026

Tipo de artículo: revisión integrativa técnico-científica, regulatoria y estratégica.

Ámbitos: industria alimentaria, química, farmacéutica y biotecnológica; Unión Europea con referencias internacionales comparadas.

VERSIÓN EDITORIAL

Resumen

La inteligencia artificial se ha convertido en un vector central de transformación de las industrias alimentaria, químico-farmacéutica y biotecnológica. Sin embargo, la ventaja competitiva no procede del algoritmo considerado aisladamente, sino de la capacidad institucional para producir, contextualizar, proteger, integrar y utilizar datos que puedan sostener inferencias científicamente válidas y decisiones defendibles. El dato industrial solo adquiere valor cuando está conectado con un fenómeno real, un método de obtención, un dominio de aplicabilidad, una decisión y una consecuencia mensurable.

Este artículo desarrolla un marco integrado para analizar el valor económico, científico, regulatorio y estratégico de los datos en sectores de alta intensidad tecnológica y elevada exposición al riesgo. Se examinan la naturaleza contextual y relacional del dato; los principios de calidad, integridad, procedencia e interoperabilidad; la aplicación de la inteligencia artificial a seguridad alimentaria, formulación, toxicología, desarrollo de medicamentos, fabricación, diseño molecular, ómicas y biología sintética; y el entramado jurídico europeo formado por el Reglamento de Inteligencia Artificial, el Reglamento de Datos, el Reglamento de Gobernanza de Datos, el RGPD, el Espacio Europeo de Datos de Salud, la protección de bases de datos y los secretos empresariales.

Se sostiene que la acumulación indiscriminada de información no constituye por sí misma capital de datos. La creación de valor exige convertir registros dispersos en productos de datos gobernados, trazables y adecuados para un contexto de uso definido. El trabajo propone un modelo de madurez, una matriz de valoración del dato, una arquitectura operativa y un itinerario de implantación orientado a riesgos y casos de uso. La conclusión principal es que, en las industrias reguladas, gobernar los datos equivale a gobernar la capacidad de innovar, demostrar cumplimiento, aprender de la experiencia y competir de forma sostenible.

Palabras clave: inteligencia artificial; gobernanza de datos; industria alimentaria; industria farmacéutica; biotecnología; industria química; integridad de datos; trazabilidad; evidencia regulatoria; innovación.

Abstract

Artificial intelligence has become a central transformation vector in the food, chemical-pharmaceutical and biotechnology industries. Competitive advantage, however, does not arise from algorithms in isolation, but from the institutional capacity to generate, contextualise, protect, integrate and use data capable of supporting scientifically valid inferences and defensible decisions. Industrial data create value only when they are linked to a real-world phenomenon, a method of generation, a domain of applicability, a decision and a measurable consequence.

This article develops an integrated framework for analysing the economic, scientific, regulatory and strategic value of data in technology-intensive, highly regulated sectors. It examines the contextual and relational nature of data; quality, integrity, provenance and interoperability requirements; artificial intelligence applications in food safety, formulation, toxicology, drug development, manufacturing, molecular design, omics and synthetic biology; and the European legal architecture formed by the Artificial Intelligence Act, Data Act, Data Governance Act, GDPR, European Health Data Space, database protection and trade-secret law.

The article argues that indiscriminate information accumulation is not equivalent to data capital. Value creation requires the transformation of fragmented records into governed, traceable data products that are fit for a defined context of use. A maturity model, a data-value assessment matrix, an operating architecture and a risk-oriented implementation roadmap are proposed. The central conclusion is that, in regulated industries, governing data is equivalent to governing the capacity to innovate, demonstrate compliance, learn from experience and compete sustainably.

Keywords: artificial intelligence; data governance; food industry; pharmaceutical industry; biotechnology; chemical industry; data integrity; traceability; regulatory evidence; innovation.

Contribución principal del artículo

El manuscrito desplaza el foco desde la adopción de herramientas de IA hacia la construcción de una capacidad empresarial verificable: transformar datos contextualizados en evidencia, decisiones y aprendizaje acumulativo. La tesis no es que “los datos sean el nuevo petróleo”, sino que constituyen un activo relacional cuyo valor depende de la calidad de los vínculos entre fenómeno, registro, modelo y acción.

Cadena de transformación del dato en valor industrial



La IA opera en el centro de la cadena; no sustituye la calidad, el contexto ni la validación.

Figura 1. Cadena de transformación del dato en valor industrial. Elaboración propia.

Aspectos destacados

- El algoritmo se está “comoditizando”; los conjuntos de datos longitudinales, contextualizados y vinculados a resultados reales siguen siendo escasos y difíciles de reproducir.
- La calidad del dato es relativa al contexto de uso. Un conjunto válido para exploración puede ser insuficiente para liberar producto, modificar un proceso o sostener una decisión regulatoria.
- Los resultados negativos, las desviaciones y los experimentos fallidos forman parte del capital científico; excluirlos produce modelos artificialmente optimistas.
- La gobernanza, la integridad, la procedencia, la interoperabilidad y la licitud no son costes externos: son componentes del valor económico y probatorio del dato.
- La implantación eficaz debe comenzar por la decisión que se pretende mejorar, no por la plataforma tecnológica que se desea adquirir.

1. Introducción: la IA no elimina la dependencia de la evidencia

La expansión de la inteligencia artificial ha generado una narrativa empresarial según la cual la disponibilidad de modelos avanzados permitiría extraer valor casi automáticamente de cualquier volumen de información. Esta narrativa es técnicamente incompleta. Un modelo puede automatizar la identificación de patrones, reducir el espacio de búsqueda experimental, optimizar parámetros o generar hipótesis; no puede reconstruir con certeza el contexto que nunca fue registrado, corregir una medición inválida, convertir una muestra sesgada en una población representativa ni resolver por sí solo la falta de derechos para utilizar un conjunto de datos.

En las industrias alimentaria, químico-farmacéutica y biotecnológica, esta limitación adquiere especial gravedad. Las salidas algorítmicas pueden influir en la seguridad de un alimento, la calidad de un medicamento, la clasificación de una sustancia, la selección de un candidato terapéutico, el diseño de una proteína o la disposición de un lote. No se trata únicamente de optimizar eficiencia. Se trata de producir conocimiento que pueda soportar consecuencias sanitarias, económicas y jurídicas.

El error conceptual más frecuente consiste en confundir digitalización con conocimiento. La digitalización convierte observaciones y actividades en registros tratables por sistemas informáticos. El

conocimiento exige, además, interpretación, contraste, contexto y capacidad para distinguir correlación, causalidad, artefacto y error. La inteligencia artificial puede intervenir de forma poderosa en esa transformación, pero continúa dependiendo de la calidad epistemológica de la infraestructura que la alimenta.

La disponibilidad comercial de modelos fundacionales, servicios en la nube y plataformas analíticas reduce las barreras de acceso al algoritmo. Por el contrario, los datos históricos de proceso, las correlaciones entre formulación y estabilidad, los patrones de contaminación de una planta, la respuesta funcional de materias primas, los resultados negativos de laboratorio, las desviaciones de fabricación y el conocimiento tácito acumulado siguen siendo activos específicos de cada organización. Esta asimetría explica por qué la ventaja sostenible se desplaza desde la mera posesión de tecnología hacia la calidad de la memoria industrial.

Desde una perspectiva económica, los datos pueden conceptualizarse como activos intangibles almacenables, no rivales y potencialmente excluibles. *Corrado, Haskel, Iommi y Jona-Lasinio (2022)* estimaron experimentalmente que la inversión en datos e inteligencia de datos representó, en seis grandes economías europeas, entre el 5 % y el 6,5 % del valor añadido bruto del sector de mercado durante 2010-2018. La estimación no implica que cualquier dato posea valor equivalente; demuestra, precisamente, que la creación de capital de datos requiere inversiones complementarias en organización, software, talento y capacidad analítica.

El objetivo de este artículo es desarrollar un marco multidisciplinar para comprender cómo se crea, conserva, destruye y gobierna el valor de los datos en tres familias industriales estrechamente relacionadas. El análisis combina economía de intangibles, ciencia de datos, tecnología alimentaria, química, farmacia, biología, gestión de calidad, propiedad intelectual y regulación europea.

2. Alcance y metodología

El trabajo adopta la forma de una revisión integrativa y crítica. No pretende realizar un metaanálisis cuantitativo de resultados algorítmicos heterogéneos, sino construir un marco conceptual y operativo a partir de cuatro tipos de fuentes: normativa europea vigente; documentos regulatorios y técnicos de organismos como EMA, FDA, EFSA, ECHA, OCDE, OMS y NIST; literatura científica revisada por pares; y principios consolidados de gestión de calidad, integridad de datos y desarrollo de modelos.

El ámbito temporal se actualiza a julio de 2026. Esta precisión es relevante porque en este momento histórico y temporal, varios instrumentos se encuentran en aplicación escalonada o en desarrollo. El Reglamento (UE) 2024/1689 de Inteligencia Artificial entró en vigor en 2024 y aplica por fases; el Reglamento de Datos es aplicable desde el 12 de septiembre de 2025; el Reglamento (UE) 2025/327 estableció el Espacio Europeo de Datos de Salud; y EMA y FDA publicaron en enero de 2026 principios conjuntos de buena práctica de IA en desarrollo de medicamentos. La guía de la FDA de enero de 2025 sobre credibilidad de modelos utilizados en decisiones regulatorias permanece, a la fecha de corte, como borrador y se trata como tal.

La comparación sectorial no presupone que los tres ámbitos sean homogéneos. La **industria alimentaria** trabaja con matrices biológicas altamente variables, cadenas de suministro extensas y sistemas preventivos como el APPCC. La **industria farmacéutica** opera bajo marcos GxP, expedientes regulatorios y requisitos explícitos de integridad. La **industria química** integra evaluación de peligros, exposición, toxicología y métodos *in silico*. La **biotecnología** añade alta dimensionalidad, datos ómicos, secuencias, estructuras y ciclos iterativos entre computación y laboratorio. La finalidad de la comparación es identificar principios comunes y diferencias que condicionan la valoración del dato.

3. Naturaleza del dato industrial: activo contextual, relacional y temporal

La metáfora del dato como “nuevo petróleo” es insuficiente para sectores científicos. El petróleo conserva propiedades físicas identificables fuera de un contexto documental. Un dato, en cambio, puede

perder casi todo su significado si se separa de la muestra, el método, el instrumento, la unidad, la fecha, el operador, la calibración, el lote, la versión del procedimiento o la población a la que pertenece.

Una lectura de pH no es plenamente interpretable sin conocer la temperatura, el método, la calibración del electrodo, la naturaleza de la matriz y el momento del proceso. Un resultado microbiológico no representa lo mismo si procede de una muestra dirigida que de un muestreo aleatorio, de una presencia/ausencia que, de un recuento, o de un método con diferente límite de detección. Una secuencia proteica adquiere un valor diferente según disponga de estructura, anotación funcional, evidencia experimental y dominio taxonómico. Un evento adverso no puede analizarse adecuadamente sin temporalidad, exposición, concomitancias y criterios de codificación.

El dato industrial es relacional porque su utilidad depende de los vínculos que permite establecer. Un resultado de estabilidad se vuelve mucho más valioso cuando puede conectarse con el lote de principio activo, el proveedor de excipiente, el equipo, el envase, la condición climática, el método analítico y la formulación exacta. La relación transforma un valor aislado en una observación explicativa.

También es temporal. Las materias primas cambian, los proveedores modifican procesos, los equipos envejecen, las poblaciones evolucionan, los métodos analíticos se actualizan y los productos se reformulan. Un modelo entrenado sobre una realidad histórica puede conservar exactitud matemática y perder relevancia operacional. La obsolescencia del dato no depende únicamente de su antigüedad, sino de la estabilidad del fenómeno representado.

Finalmente, el dato es institucional. Su valor depende de que una organización pueda localizarlo, comprenderlo, utilizarlo legalmente, reproducir las transformaciones realizadas y asignar responsabilidad sobre su mantenimiento. Los registros dispersos en hojas de cálculo personales, correos electrónicos, equipos aislados o documentos escaneados existen, pero no constituyen todavía un activo plenamente activable.

Propiedad	Pregunta de control	Consecuencia sobre el valor
Procedencia	¿Quién, cómo, cuándo y con qué sistema generó el dato?	Determina credibilidad, reproducibilidad y valor probatorio.
Contexto	¿Qué matriz, población, proceso y condiciones representa?	Delimita interpretación y dominio de aplicabilidad.
Semántica	¿La variable mantiene significado, unidad y codificación comunes?	Permite integración y evita comparaciones falsas.
Temporalidad	¿Sigue representando la realidad actual?	Condiciona deriva, revalidación y uso histórico.
Licitud (Concepto)	¿Qué base jurídica, contrato, licencia o secreto limita su uso?	Define si el valor potencial puede convertirse en valor realizable.
Accionable (Concepto)	¿Puede modificar una decisión o reducir una incertidumbre?	Diferencia el dato almacenado del dato económicamente útil.

Tabla 1. Propiedades que convierten el registro en activo de datos. Elaboración propia.

4. De los datos al capital de datos

La mera acumulación de registros no equivale a capital. El capital de datos aparece cuando la información puede reutilizarse de manera controlada para generar beneficios futuros: reducir incertidumbre, acelerar experimentos, anticipar fallos, sostener evidencia, optimizar recursos o mejorar decisiones. Esta definición exige diferenciar cuatro niveles.

El primer nivel es el dato bruto: mediciones, eventos o documentos conservados con mínima elaboración. El segundo es el dato estructurado y contextualizado: el registro se vincula con metadatos,

identificadores y reglas semánticas. El tercero es el producto de datos: un conjunto mantenido, documentado, accesible y preparado para usuarios y finalidades definidas. El cuarto es el activo de conocimiento: el producto de datos se integra con modelos, experiencia y reglas de decisión y produce resultados repetibles.

La no rivalidad del dato permite que un mismo registro genere usos múltiples. Un espectro puede servir para liberar materia prima, detectar adulteración, estudiar variabilidad de proveedores y entrenar un modelo de clasificación. Esta reutilización explica los rendimientos crecientes potenciales. Sin embargo, el valor marginal no crece indefinidamente. Los registros redundantes, carentes de variabilidad o desconectados de resultados pueden añadir costes de almacenamiento, seguridad y revisión sin aportar información nueva.

El valor también depende de la complementariedad. Un conjunto de datos excelente puede permanecer estéril sin capacidad analítica, conocimiento sectorial, infraestructura, cultura de uso y autoridad para cambiar procesos. Por eso, la contabilidad tradicional subestima el activo y, al mismo tiempo, el discurso tecnológico suele sobreestimarlos: la primera no registra adecuadamente la inversión; el segundo atribuye valor a información que todavía no puede ser utilizada.

4.1. Un índice de valor del dato para la gestión industrial

Para priorizar inversiones puede emplearse un Índice de Valor del Dato (IVD) como instrumento directivo. No se presenta como una métrica científica universal ni como una valoración contable, sino como una heurística auditable. Cada producto de datos se puntúa de 0 a 5 en ocho dimensiones: relevancia decisoria, calidad, singularidad, cobertura, interoperabilidad, trazabilidad, disponibilidad jurídica y actualidad. Las ponderaciones deben definirse según el caso de uso.

$$\text{IVD} = \sum (w_i \times s_i), \text{ con } \sum w_i = 1$$

Instrumento propuesto de priorización; no constituye una métrica validada externamente.

Dimensión	0–1: valor bajo	2–3: valor intermedio	4–5: valor alto
Relevancia decisoria	No modifica decisiones.	Apoya análisis, pero no activa acciones claras.	Cambia decisiones críticas o reduce riesgos materiales.
Calidad	Errores, ausencias o sesgo no controlados.	Calidad suficiente con limitaciones conocidas.	Controles robustos, incertidumbre y representatividad documentadas.
Singularidad	Fácilmente disponible o reproducible.	Requiere cierta inversión o acceso.	Histórico propio, raro, longitudinal o costoso de reproducir.
Interoperabilidad	Formato cerrado y semántica local.	Conversión posible con trabajo manual.	Estándares, ontologías, API o contratos de datos estables.
Trazabilidad	Origen o transformaciones desconocidos.	Linaje parcial.	Procedencia y transformaciones completamente reproducibles.
Disponibilidad jurídica	Uso incierto o prohibido.	Uso limitado por contrato/finalidad.	Derechos, licencias y restricciones claramente gobernados.
Actualidad	No representa el proceso vigente.	Requiere ajustes o reponderación.	Representa la realidad operacional actual.
Cobertura	Casos críticos ausentes.	Cobertura desigual.	Incluye variabilidad normal, extremos y eventos relevantes.

Tabla 2. Matriz propuesta para estimar el valor estratégico de un producto de datos.

5. Calidad, integridad, procedencia e interoperabilidad

La calidad del dato no es una propiedad absoluta. Debe evaluarse respecto a una finalidad. Un muestreo puede ser suficiente para monitorizar una tendencia y resultar insuficiente para demostrar ausencia de un peligro. Datos de sensores agregados cada minuto pueden servir para eficiencia energética y ocultar una transición crítica de segundos. Una base de reclamaciones puede detectar patrones, pero no estimar prevalencia cuando la propensión a reclamar varía entre clientes.

El marco METRIC, desarrollado para inteligencia artificial médica, identifica dimensiones como representatividad, completitud, corrección, actualidad, diversidad, consistencia, relevancia y documentación. Su principio esencial es transferible: el uso previsto y el modelo condicionan la evaluación de la aptitud del conjunto (*Schwabe et al.*, 2024). En regulación de medicamentos, EMA ha desarrollado un *Data Quality Framework* que vincula calidad con propósito, evidencia y confianza regulatoria (EMA, 2023).

En entornos GxP, la integridad exige que los datos sean atribuibles, legibles, contemporáneos, originales y exactos, junto con atributos de completitud, consistencia, perdurabilidad y disponibilidad. Estos principios ALCOA+ no deben reducirse a una lista formal. La atribución exige conocer las intervenciones humanas y automáticas; la originalidad requiere preservar datos brutos y metadatos; la exactitud incluye calibración y control del sistema; la completitud alcanza registros de auditoría, exclusiones y repeticiones.

La procedencia o linaje describe el recorrido desde la fuente hasta la salida. Debe permitir reconstruir adquisiciones, conversiones, filtros, imputaciones, normalizaciones, etiquetados y agregaciones. En modelos de IA, la procedencia también alcanza la versión del conjunto, el código, las bibliotecas, los hiperparámetros, las semillas aleatorias y el entorno de ejecución.

Los principios FAIR —hallables, accesibles, interoperables y reutilizables— proporcionan una orientación útil para la ciencia de datos. FAIR no significa necesariamente abierto. Los datos pueden ser accesibles mediante autenticación y autorización. Tampoco garantiza calidad intrínseca, privacidad o ética. Su aportación consiste en hacer que datos y metadatos puedan localizarse, comprenderse e integrarse bajo reglas explícitas (*Wilkinson et al.*, 2016).

Dimensión	Control mínimo	Fallo típico en IA	Medida correctora
Representatividad	Comparar conjunto y población objetivo.	Buen rendimiento medio, fallo en materias primas, plantas o subgrupos no representados.	Muestreo estratificado, validación externa y límites de uso.
Exactitud metrológica	Calibración, incertidumbre y método.	El modelo aprende ruido o diferencias instrumentales.	Armonización, materiales de referencia y transferencia de calibración.
Completitud contextual	Metadatos obligatorios.	Correlaciones espurias por variables omitidas.	Diccionario, captura automática y reglas de rechazo.
Consistencia semántica	Unidades y vocabularios controlados.	Mezcla de códigos, métodos o definiciones.	Ontologías, MDM y contratos de datos.
Linaje	Registro de transformaciones y versiones.	Resultado no reproducible ni auditable.	Pipelines versionados y registro inmutable de ejecución.
Actualidad	Monitorizar cambios de proceso/población.	Deriva silenciosa.	Indicadores de deriva y eventos de revalidación.
Independencia	Separar entrenamiento, prueba y confirmación.	Validación circular y sobreajuste.	Conjunto bloqueado, validación externa o prospectiva.
Incetidumbre	Conservar error y límites de detección.	Predicciones excesivamente precisas.	Intervalos, calibración probabilística y abstención.

Tabla 3. Dimensiones de calidad y controles para sistemas industriales de IA.

6. Industria alimentaria: convertir trazabilidad en capacidad predictiva

La industria alimentaria produce datos desde la explotación primaria hasta el consumidor: origen, certificaciones, composición, alérgenos, contaminantes, condiciones de transporte, parámetros de proceso, limpieza, ambiente, microbiología, análisis sensorial, vida útil, reclamaciones y devoluciones. Pese a esta abundancia, la información suele permanecer fragmentada entre ERP, LIMS, MES, registros APPCC, plataformas de proveedores, sensores y documentos de calidad.

La trazabilidad exigida por el Reglamento (CE) N.º 178/2002 permite identificar el movimiento de alimentos, piensos, animales e ingredientes. Para la IA industrial resulta necesario ampliar esa lógica hacia una trazabilidad causal: relacionar cada transformación con condiciones, propiedades y resultados. Saber de qué proveedor procede una materia prima es necesario; conocer qué atributos funcionales, microbiológicos y fisicoquímicos presentaba y cómo interactuó con el proceso es lo que permite aprender.

La variabilidad biológica es un desafío estructural. Ingredientes con especificaciones equivalentes pueden comportarse de manera distinta por variedad, campaña, origen, tratamiento térmico, granulometría, desnaturalización, humedad o historia de almacenamiento. Si la base solo conserva la denominación comercial y el porcentaje de formulación, el modelo atribuirá a otros factores una variación que procede de propiedades no registradas.

6.1. Seguridad alimentaria y APPCC aumentado

La inteligencia artificial puede apoyar detección de contaminantes, predicción microbiológica, inspección visual, priorización de muestreos, mantenimiento higiénico, vigilancia de señales y análisis de trazabilidad. EFSA ha situado la preparación de datos para IA y la integración de grandes volúmenes de información entre sus prioridades recientes. La Comisión Europea utiliza sistemas de información como IMSOC y TRACES, y herramientas de extracción de información como *TraceMap*, para facilitar el análisis de redes de fraude e incidentes.

Estas capacidades no sustituyen los programas de prerequisites ni el APPCC. Un modelo no compensa una zonificación higiénica deficiente, una limpieza no validada, un plan de muestreo inadecuado o una cultura de seguridad débil. *Motzer et al. (2024)*, al estudiar agricultura en ambiente controlado, advierten precisamente que implantar soluciones intensivas en datos sobre sistemas de seguridad inmaduros puede consolidar un punto de partida inapropiado.

Un APPCC aumentado por IA debe conservar la lógica preventiva: peligro definido, medida de control, criterio de actuación, responsable y registro. La alerta predictiva ha de integrarse en procedimientos que establezcan qué ocurre ante un positivo, un resultado incierto, una contradicción con el criterio experto o una indisponibilidad del sistema. La supervisión humana solo es real cuando el responsable dispone de tiempo, información, competencia y autoridad para apartarse de la recomendación.

Los errores deben valorarse según su coste. En detección de peligros graves puede priorizarse sensibilidad, aceptando más falsos positivos. En control de calidad visual, una tasa excesiva de falsos rechazos puede hacer inviable el sistema. La métrica adecuada no es la exactitud global, sino la función de pérdida asociada a las consecuencias reales.

6.2. Formulación, proceso, sensorialidad y vida útil

En desarrollo de producto, la IA puede relacionar composición, propiedades de ingredientes, proceso, estructura, sensorialidad, estabilidad, coste y aceptación. La dificultad principal no es seleccionar un algoritmo, sino construir un conjunto que refleje el espacio de formulación.

Los datos históricos presentan sesgo de supervivencia. Las formulaciones exitosas se documentan con detalle; los ensayos descartados se archivan de forma irregular. El modelo aprende así de una población enriquecida artificialmente y no conoce adecuadamente los límites del sistema. Registrar fracasos, motivos

de descarte y condiciones experimentales reduce repeticiones y mejora la búsqueda bayesiana, el aprendizaje activo y el diseño de experimentos.

La escala introduce otra discontinuidad. Una formulación de laboratorio no se comporta necesariamente igual en piloto o planta. Geometría, cizalla, transferencia térmica, tiempo de residencia y secuencia de adición pueden modificar estructura y funcionalidad. Un gemelo digital útil debe integrar conocimiento mecanístico y datos multi-escala; limitarse a extrapolar correlaciones de laboratorio puede producir una falsa confianza.

En vida útil, los modelos deben distinguir entre predicción y demostración. Pueden estimar la evolución de calidad o riesgo, seleccionar condiciones y optimizar diseños de estudio. La justificación final exige ensayos adecuados, hipótesis explícitas, incertidumbre y validación en condiciones representativas. La IA reduce experimentos innecesarios; no convierte datos insuficientes en una prueba concluyente.

6.3. Cadena de suministro, autenticidad y sostenibilidad

La combinación de sensores, espectroscopía, visión, geolocalización, datos comerciales y registros de certificación puede reforzar autenticidad y trazabilidad. Los modelos pueden detectar patrones incompatibles con origen, composición o comportamiento esperado. Sin embargo, una cadena digital no es necesariamente una cadena veraz: un sistema distribuido preserva la inmutabilidad de lo registrado, no garantiza que el dato de entrada sea correcto.

La interoperabilidad entre empresas es crítica. La falta de identificadores, vocabularios y formatos comunes obliga a conversiones manuales y dificulta la trazabilidad extremo a extremo. El valor de los espacios de datos agroalimentarios dependerá de su capacidad para establecer reglas de confianza, derechos de uso, modelos semánticos y mecanismos de calidad, no únicamente de conectar plataformas.

Caso de uso alimentario	Datos necesarios	Valor potencial	Riesgo crítico
Predicción de contaminación ambiental	Muestreos, zonas, producción, limpieza, flujo de personal, temperatura, mantenimiento.	Muestreo adaptativo y detección temprana.	Eventos raros, sesgo de muestreo y falsa sensación de control.
Vida útil	Formulación, proceso, envase, temperatura, microbiología, sensorial, pH, aw.	Diseños de estudio más eficientes y reducción de desperdicio.	Extrapolación fuera de condiciones validadas.
Formulación asistida	Propiedades funcionales, composición, proceso, coste, estabilidad, sensorial.	Reducción del espacio experimental y optimización multicriterio.	Ausencia de datos negativos y variables latentes.
Inspección visual/espectral	Imágenes o espectros etiquetados, calibración, iluminación, lotes y referencia.	Control en línea, autenticidad y clasificación.	Cambio de instrumento, temporada o presentación.
Trazabilidad y fraude	Lotes, transacciones, origen, certificados, rutas, análisis.	Investigación rápida y detección de anomalías.	Datos de entrada falsos o semántica incompatible.

Tabla 4. Casos de uso de IA y condiciones de datos en la industria alimentaria.

7. Industria químico-farmacéutica: el dato como evidencia regulatoria

La industria químico-farmacéutica combina datos moleculares, toxicológicos, analíticos, preclínicos, clínicos, productivos y poscomercialización. La IA puede utilizarse en descubrimiento de dianas, diseño molecular, predicción de propiedades, optimización de síntesis, selección de pacientes, análisis de imágenes, farmacovigilancia, control de proceso y evaluación de datos del mundo real.

La característica diferencial es que una salida puede integrarse en evidencia regulatoria. Esto eleva la exigencia sobre la credibilidad del modelo y obliga a definir con precisión el contexto de uso. Un modelo empleado para generar hipótesis internas no requiere el mismo nivel de demostración que uno utilizado para

establecer un atributo crítico, justificar una estrategia de control o producir información que respalde seguridad, eficacia o calidad.

El documento de reflexión de EMA de 2024 subraya un enfoque humano, proporcional al riesgo y extendido a todo el ciclo de vida del medicamento. La guía en borrador de la FDA de 2025 propone un marco de credibilidad basado en riesgo para modelos que produzcan información destinada a decisiones regulatorias. En enero de 2026, EMA y FDA publicaron diez principios conjuntos: diseño centrado en la persona, enfoque basado en riesgo, normas, contexto de uso claro, experiencia multidisciplinar, gobernanza y documentación de datos, buenas prácticas de diseño, evaluación del rendimiento, gestión del ciclo de vida e información esencial clara.

La convergencia regulatoria es significativa: la aceptabilidad no depende de la sofisticación del modelo, sino de que la pregunta, los datos, la incertidumbre, el rendimiento, los límites y la supervisión sean comprensibles y proporcionales a la consecuencia.

7.1. Descubrimiento y desarrollo de medicamentos

Los modelos pueden priorizar dianas, predecir afinidad, propiedades ADME/Tox, sintetizabilidad y actividad. Su valor económico reside en reducir el número de candidatos y experimentos antes de comprometer recursos mayores. El criterio no debe ser cuántas moléculas genera el sistema, sino cuánto mejora la probabilidad de seleccionar compuestos viables y cuánto reduce el tiempo hasta una decisión de continuar o detener.

La validación retrospectiva sobre bases públicas puede sobreestimar el rendimiento. Los conjuntos contienen sesgos de publicación, duplicados, diferencias de protocolo y mediciones no comparables. Las particiones aleatorias permiten que moléculas muy similares aparezcan en entrenamiento y prueba. Las divisiones temporales, por estructura o por serie química se aproximan mejor al desafío prospectivo.

Los “*activity cliffs*” muestran que pequeñas modificaciones estructurales pueden causar grandes cambios de actividad y desafiar modelos QSAR. El dominio de aplicabilidad debe documentarse. Una predicción fuera de dominio no debería presentarse como equivalente a una observación experimental.

7.2. Fabricación, calidad y liberación

En fabricación farmacéutica, la IA puede analizar parámetros críticos, atributos de calidad, datos PAT, tendencias de laboratorio, estados de equipo y desviaciones. La oportunidad consiste en detectar señales antes de que se conviertan en fallos y mejorar la comprensión del proceso. El riesgo consiste en implantar una capa opaca paralela al sistema de calidad.

Cuando una salida influye en una decisión GMP, deben controlarse datos de entrada, acceso, cambios, auditoría, versión y recuperación. El modelo forma parte del sistema informatizado y su criticidad depende de su función. La validación debe demostrar que el sistema es apto para el uso previsto, no que alcanza una métrica genérica.

Los datos de fabricación comercial se concentran alrededor de condiciones controladas. Por diseño contienen pocos fallos críticos. Esta escasez de eventos impide confiar exclusivamente en aprendizaje supervisado. Es necesario combinar conocimiento de proceso, estudios de caracterización, simulación, datos de desarrollo, gestión de riesgos de calidad y evaluación experta.

La liberación en tiempo real o la predicción de atributos no deben confundirse con una simple inferencia matemática. Exigen una estrategia analítica y de control validada, gestión de incertidumbre, procedimientos de contingencia y criterios para el comportamiento fuera de distribución.

7.3. Datos clínicos, evidencia del mundo real y farmacovigilancia

Los datos clínicos y del mundo real amplían la capacidad para estudiar efectividad, seguridad, uso y subpoblaciones. También presentan ausencias, errores de codificación, confusión, sesgo de selección y

cambios en la práctica asistencial. La cantidad no elimina estos problemas; puede hacer que estimaciones sesgadas parezcan extremadamente precisas.

La guía EMA sobre sistemas informatizados y datos electrónicos en ensayos clínicos exige gobernanza durante todo el ciclo de vida. ICH E6(R3) refuerza el enfoque proporcionado al riesgo, la calidad por diseño y la fiabilidad de registros. La IA empleada en selección, adjudicación, monitorización o análisis debe integrarse en el diseño del estudio y documentar cómo afecta a derechos, seguridad y fiabilidad de resultados.

En farmacovigilancia, el procesamiento de lenguaje puede priorizar casos y extraer entidades, pero las decisiones de causalidad y señal requieren juicio. Los modelos entrenados con prácticas históricas pueden reproducir inconsistencias de codificación y sub-representar eventos raros. Deben evaluarse sensibilidad, valor predictivo, carga de revisión y posibles efectos sobre grupos específicos.

7.4. Industria química, REACH y métodos *in silico*

La industria química utiliza datos de identidad, propiedades fisicoquímicas, toxicidad, ecotoxicidad, exposición y usos. REACH promueve la puesta en común de determinados datos y la evitación de ensayos innecesarios con animales. Los modelos (Q)SAR y el *read-across* constituyen precedentes regulatorios directos de inferencia basada en datos.

ECHA exige que un (Q)SAR esté científicamente validado, que la sustancia se encuentre dentro de su dominio de aplicabilidad y que la predicción sea adecuada para la finalidad. La OCDE ha establecido principios y marcos de evaluación que incluyen *endpoint* definido, algoritmo no ambiguo, dominio de aplicabilidad, medidas de ajuste, robustez e índices de predictibilidad, y, cuando sea posible, interpretación mecanística.

La lección para la IA contemporánea es contundente: la complejidad algorítmica no sustituye una justificación científica. Un modelo más opaco no es automáticamente más convincente. En contextos regulatorios, la transparencia puede exigir explicaciones sobre datos, dominio, incertidumbre y plausibilidad, incluso cuando no sea posible interpretar cada parámetro interno.

Contexto	Nivel de impacto	Evidencia esperable	Ejemplo de control
Exploración interna	Bajo o indirecto	Rendimiento técnico, documentación suficiente para reproducir y uso limitado.	Revisión experta y etiquetado “no regulatorio”.
Priorización experimental	Medio	Validación retrospectiva rigurosa, prueba prospectiva y dominio de aplicabilidad.	Criterios de abstención y confirmación experimental.
Apoyo a estrategia de desarrollo	Medio-alto	Contexto de uso, análisis de riesgo, robustez y sensibilidad.	Plan de credibilidad y gestión de cambios.
Información para decisión regulatoria	Alto	Evidencia proporcional a la consecuencia, documentación completa y diálogo regulatorio.	Validación independiente y trazabilidad del ciclo de vida.
Control o liberación de fabricación	Alto/crítico	Integración GxP, validación del sistema, contingencia y monitorización.	Procedimientos de fallo seguro y revalidación.

Tabla 5. Proporcionalidad entre contexto de uso y credibilidad exigible.

8. Biotecnología: datos de alta dimensionalidad y ciclos cerrados de aprendizaje

La biotecnología integra secuencias, estructuras, perfiles ómicos, imágenes, fenotipos, datos de cultivo, fermentación y bioensayo. La IA puede proponer proteínas, enzimas, anticuerpos, rutas metabólicas,

condiciones de proceso y biomarcadores. En este sector, la diferencia entre predicción y función experimental es especialmente visible.

Los modelos de diseño proteico combinan tres modalidades principales: secuencia, estructura y etiquetas funcionales (Notin *et al.*, 2024). La disponibilidad de grandes bases de secuencias y estructuras ha impulsado avances extraordinarios, pero la anotación funcional continúa siendo más escasa, heterogénea y costosa. Predecir una estructura no demuestra actividad, selectividad, estabilidad, expresión, inmunogenicidad ni comportamiento en una matriz industrial.

El laboratorio sigue siendo el mecanismo de validación. La IA puede seleccionar experimentos de mayor valor informativo, y la automatización puede cerrar ciclos diseño-construcción-prueba-aprendizaje. Para que el ciclo sea acumulativo, cada ejecución debe generar datos normalizados, metadatos completos y resultados positivos y negativos.

8.1. Resultados negativos, sesgo de publicación y memoria experimental

Los resultados negativos delimitan el espacio de posibilidades. Permiten descartar regiones, identificar incompatibilidades y evitar repeticiones. Sin embargo, suelen documentarse peor que los éxitos. Esta asimetría produce un sesgo interno equivalente al sesgo de publicación científica.

Un modelo entrenado solo con candidatos seleccionados aprende una realidad artificial. No puede estimar adecuadamente la probabilidad de fracaso y puede reproducir estrategias ya exploradas. La empresa que estructura su historial de intentos fallidos crea una cartografía privada de alto valor, difícil de adquirir en el mercado.

La documentación debe distinguir ausencia de efecto, fallo técnico, ensayo inconcluso, muestra inválida y resultado contradictorio. Mezclar estas categorías degrada el aprendizaje. Un “negativo” por debajo del límite de detección no equivale a una actividad demostrada como inexistente.

8.2. Biocuración, ontologías y conocimiento legible por máquinas

Las bases biológicas requieren curación. La biocuración transforma artículos, experimentos y anotaciones en conocimiento estructurado y legible por máquinas. Dessimoz y Thomas (2024) sostienen que la revolución de IA aumenta, no reduce, la necesidad de curación experta.

Las ontologías y grafos de conocimiento permiten representar entidades, relaciones y niveles de evidencia. Deben distinguir evidencia directa, inferencia computacional, anotación automática, resultado preliminar y consenso. Esta diferenciación evita que el sistema trate todas las afirmaciones como equivalentes.

La organización que combina datos internos con literatura, patentes y bases públicas necesita conservar procedencia y licencia. La integración semántica no debe borrar el origen. Una inferencia generada por un modelo no puede reingresar en la base como si fuera una observación confirmada.

8.3. Datos sintéticos y aumentación

Los datos sintéticos pueden aumentar conjuntos escasos, proteger parcialmente la privacidad, simular escenarios y equilibrar clases. No son automáticamente anónimos ni fieles. Pueden memorizar individuos, reproducir sesgos, suavizar eventos raros o generar combinaciones biológicamente imposibles (van Breugel *et al.*, 2024).

Deben identificarse y versionarse separadamente. La validación ha de evaluar utilidad, fidelidad, diversidad, privacidad y riesgo de inferencia. No es metodológicamente sólido generar datos con un modelo y utilizarlos como confirmación independiente del mismo supuesto sin contraste externo.

Activo biotecnológico	Metadatos críticos	Valor de IA	Riesgo de pérdida de contexto
Secuencias	Organismo, constructo, versión, cobertura, calidad.	Diseño, predicción de variantes y anotación.	Errores de ensamblaje o procedencia taxonómica.

Activo biotecnológico	Metadatos críticos	Valor de IA	Riesgo de pérdida de contexto
Estructuras	Método, resolución, conformación, ligando, condiciones.	Predicción de interacción y diseño.	Tratar una conformación como estado único.
Ómicas	Preparación, lote, plataforma, normalización, covariables.	Biomarcadores, estratificación y redes.	Efectos de lote y alta dimensionalidad.
Bioensayos	Protocolo, controles, concentración, tiempo, réplicas.	Aprendizaje activo y selección de candidatos.	Etiquetas inconsistentes y negativos inválidos.
Fermentación/bioproseso	Cepa, pasaje, medio, alimentación, sensores, escala.	Control, optimización y gemelo digital.	Deriva biológica y transferencia de escala.
Imágenes celulares	Microscopio, óptica, tinción, exposición, segmentación.	Fenotipado y control de calidad.	Atajos visuales y diferencias de equipo.

Tabla 6. Activos de datos característicos de la biotecnología.



Figura 2. Diferencias sectoriales y núcleo común de creación de valor. Elaboración propia.

9. Arquitectura jurídica europea del dato y la IA

No existe un derecho general y unitario de propiedad sobre cualquier dato. El control puede derivar de protección de datos personales, contratos, secretos empresariales, derechos sobre bases de datos, propiedad intelectual sobre contenidos, deberes de confidencialidad, normativa sectorial o control técnico de acceso. Por ello, el hecho de afirmar que una empresa “es propietaria de los datos” puede ocultar una cadena de derechos y restricciones más compleja.

La Directiva 96/9/CE protege la estructura original de determinadas bases mediante derechos de autor y reconoce, cuando se cumplen los requisitos, un derecho sui generis vinculado a la inversión en obtención, verificación o presentación. La Directiva (UE) 2016/943 protege información secreta con valor empresarial cuando ha sido objeto de medidas razonables de protección. La razonabilidad de las medidas exige clasificación, acceso, contratos, seguridad y respuesta frente a extracción.

El RGPD se aplica cuando existe información sobre personas identificadas o identificables. Los datos de salud, genéticos y biométricos reciben protección reforzada. La seudonimización reduce riesgo, pero no equivale a anonimización. El Comité Europeo de Protección de Datos ha señalado que la anonimidad de un modelo entrenado con datos personales debe evaluarse caso por caso, considerando extracción, inferencia y medios razonablemente disponibles.

El Reglamento de Gobernanza de Datos regula mecanismos destinados a aumentar la confianza en la reutilización e intermediación y apoya espacios comunes europeos. El Reglamento de Datos establece reglas de acceso y uso, especialmente sobre datos generados por productos conectados y servicios relacionados, y es aplicable desde septiembre de 2025. Para maquinaria industrial y equipos analíticos, el contrato debe analizar quién accede a datos, en qué formato, con qué latencia y con qué derechos de reutilización.

El Reglamento (UE) 2025/327 creó el Espacio Europeo de Datos de Salud, con reglas para uso primario y secundario. Su implementación será gradual, pero altera el marco estratégico para investigación, farmacovigilancia y evidencia del mundo real.

El Reglamento de Inteligencia Artificial adopta un enfoque basado en riesgo. La pertenencia a una industria regulada no convierte automáticamente cualquier sistema en alto riesgo. La clasificación depende de función, contexto, categorías del Reglamento y, en su caso, integración en productos regulados. Un sistema no clasificado como alto riesgo continúa sujeto a legislación de producto, seguridad, datos, contratos, propiedad intelectual y responsabilidad.

Instrumento	Objeto principal	Implicación empresarial para datos e IA
Reglamento (UE) 2024/1689 · AI Act	Sistemas de IA y modelos de propósito general; obligaciones basadas en riesgo.	Inventario de sistemas, clasificación, alfabetización, documentación, transparencia y gobierno del ciclo de vida según rol y uso.
Reglamento (UE) 2023/2854 · Data Act	Acceso y utilización justa de datos, especialmente de productos conectados.	Revisar contratos de maquinaria, sensores, laboratorios y servicios cloud; portabilidad y acceso.
Reglamento (UE) 2022/868 · Data Governance Act	Intermediación, reutilización y confianza en el intercambio.	Participación en espacios de datos bajo gobernanza y condiciones de acceso.
Reglamento (UE) 2016/679 · RGPD	Protección de datos personales.	Base jurídica, minimización, seguridad, derechos, evaluaciones de impacto y transferencias.
Reglamento (UE) 2025/327 · EHDS	Uso primario y secundario de datos electrónicos de salud.	Nuevas infraestructuras y requisitos para investigación y evidencia sanitaria.
Directiva 96/9/CE	Protección jurídica de bases de datos.	Evaluar originalidad, inversión y límites de extracción/reutilización.
Directiva (UE) 2016/943	Secretos empresariales.	Implantar medidas razonables de protección y controlar uso de IA externa.
Reglamento (CE) 178/2002	Principios generales y trazabilidad alimentaria.	Conectar datos de trazabilidad con seguridad y responsabilidad del operador.
Reglamento (CE) 1907/2006 · REACH	Registro y evaluación de sustancias químicas.	Gobierno de datos toxicológicos, puesta en común y justificación de métodos alternativos.

Tabla 7. Mapa jurídico esencial. Debe completarse para cada producto, territorio y contexto de uso.

9.1. Contratos, nube y pérdida de control

La contratación tecnológica es una parte de la gobernanza del dato. Deben revisarse finalidad, localización, sub-encargados, retención, uso para entrenamiento, confidencialidad, propiedad de salidas, portabilidad, registros, seguridad, auditoría, continuidad y terminación.

La cláusula que reconoce al cliente titularidad sobre sus datos puede ser insuficiente si el proveedor controla el formato, los metadatos, las interfaces o la posibilidad práctica de exportación. El bloqueo no es solo jurídico; puede ser semántico y operativo.

El uso de sistemas generativos públicos con formulaciones, secuencias, informes de desviación, resultados clínicos o hipótesis patentables puede comprometer confidencialidad y secreto. La política debe definir categorías de información, herramientas autorizadas, configuración, revisión humana y tratamiento de salidas.

9.2. Compartición controlada y aprendizaje federado

La innovación requiere combinar datos entre actores, pero la concentración no siempre es necesaria. El aprendizaje federado permite entrenar modelos distribuidos y puede reducir transferencias directas. No elimina riesgos: actualizaciones pueden filtrar información, participantes pueden introducir datos maliciosos y la heterogeneidad local puede degradar rendimiento.

Durrant et al. (2022) analizaron el valor del aprendizaje federado entre silos en el sector agroalimentario; *Gavai et al.* (2023) exploraron su aplicación al fraude alimentario. Estas arquitecturas son prometedoras cuando se acompañan de contratos, seguridad, agregación, evaluación de contribuciones, control de versiones y gobernanza común.

10. Arquitectura empresarial para convertir datos en evidencia

Una estrategia madura no comienza construyendo un gran lago de datos. Comienza definiendo decisiones y riesgos. Los repositorios centralizados sin semántica ni responsabilidad pueden convertirse en pantanos de datos: costosos, duplicados y poco confiables.

La unidad operativa recomendable es el producto de datos. Se trata de un conjunto mantenido para una finalidad, con propietario, usuarios, definición, calidad, derechos, linaje y niveles de servicio. Ejemplos son un producto de datos de vida útil, desviaciones, estabilidad, expresión proteica, control ambiental o reclamaciones.

Cada producto debe declarar: finalidad; propietario funcional; responsable técnico; fuentes; modelo semántico; controles; reglas de acceso; base jurídica; retención; frecuencia; indicadores; procedimiento de corrección; dependencia de terceros; y criterios de retirada.

La arquitectura debe preservar continuidad entre realidad física, captura, gobierno, modelización, validación y decisión. Cuando estas capas se gestionan como proyectos separados, la trazabilidad se rompe.

Arquitectura de datos e IA para industrias reguladas



El valor depende de la continuidad entre el fenómeno real, el registro y la decisión.

Figura 3. Arquitectura de datos e IA para industrias reguladas. Elaboración propia.

10.1. Modelo de responsabilidades

El gobierno no puede delegarse exclusivamente en tecnología. La dirección define apetito de riesgo y prioridades. El propietario del proceso responde de la decisión. El propietario del dato define significado y calidad. El responsable de plataforma asegura disponibilidad y controles. El equipo científico establece plausibilidad y diseño. Calidad y regulación evalúan impacto y evidencia. Jurídico y privacidad delimitan derechos. Ciberseguridad protege confidencialidad, integridad y disponibilidad.

En sistemas críticos debe existir independencia suficiente entre desarrollo y validación. No siempre requiere equipos totalmente separados, pero sí revisión objetiva, criterios predefinidos y autoridad para rechazar el despliegue.

Rol	Responsabilidad no delegable	Evidencia esperada
Dirección/Consejo	Prioridad, recursos y apetito de riesgo.	Política, cartera y revisión de riesgos.
Propietario del proceso	Resultado operativo y decisión.	Procedimiento, KPI y aceptación del sistema.
Propietario del dato	Definición, calidad, acceso y ciclo de vida.	Ficha del producto de datos y controles.
Ciencia/tecnología	Plausibilidad, método, incertidumbre y límites.	Protocolo, informe de desarrollo y análisis.
Calidad/Regulatorio	Adecuación al sistema de calidad y uso regulatorio.	Evaluación de impacto, validación y change control.
Legal/Privacidad	Derechos, base jurídica, contratos y secreto.	Registro de tratamiento, DPIA, licencias y cláusulas.
Ciberseguridad	Amenazas, accesos, continuidad y respuesta.	Modelo de amenazas, registros y pruebas.
Usuario experto	Uso conforme, revisión y escalado.	Formación, registro de decisiones e incidencias.

Tabla 8. Responsabilidades mínimas en el modelo operativo de datos e IA.

11. Gobierno del ciclo de vida del modelo

La validación no es una certificación permanente. Un modelo es aceptable para un contexto de uso, una población, una versión de datos y unas condiciones. El ciclo debe comenzar con la pregunta de interés y terminar con monitorización, retirada y conservación de evidencia.

El expediente debe permitir reconstruir problema, decisión, usuarios, datos, transformaciones, código, arquitectura, métricas, criterios, resultados, limitaciones, supervisión, seguridad y cambios. Debe diferenciar rendimiento técnico, utilidad operativa y beneficio empresarial.

La evaluación debe incorporar comparación con la práctica existente. Una mejora estadística pequeña puede no justificar coste, complejidad o riesgo. Un modelo menos preciso pero interpretable, estable y fácil de integrar puede generar más valor.

Debe existir capacidad de abstención. Cuando la entrada se encuentra fuera del dominio o la incertidumbre supera un umbral, el sistema debe remitir a revisión o utilizar una estrategia segura. Obligar al modelo a responder siempre convierte desconocimiento en falsa precisión.



La validación no es un acto único: es una obligación continua vinculada al contexto de uso.

Figura 4. Ciclo de vida de un sistema de IA industrial. Elaboración propia.

Fase	Preguntas de control	Salida documental
Problema	¿Qué decisión mejora? ¿Qué daño puede causar? ¿Existe alternativa más simple?	Caso de uso, propietario y evaluación preliminar de riesgo.
Datos	¿Representan el contexto? ¿Son lícitos, trazables y suficientes?	Ficha de datos, calidad, derechos y limitaciones.
Desarrollo	¿Qué método y por qué? ¿Cómo se evitan fuga y sobreajuste?	Plan de desarrollo, código versionado y experimentos.
Validación	¿Cumple criterios? ¿En qué subgrupos falla? ¿Está calibrado?	Informe independiente y declaración de uso.

Fase	Preguntas de control	Salida documental
Despliegue	¿Cómo se integra? ¿Quién revisa? ¿Qué ocurre si falla?	Procedimiento, formación, controles y contingencia.
Vigilancia	¿Hay deriva, incidentes o cambios? ¿Cuándo revalidar?	Panel, registro de incidentes y change control.
Retirada	¿Cómo se desactiva y conserva evidencia?	Plan de retirada, archivo y sustitución.

Tabla 9. Entregables del ciclo de vida del modelo.

12. Riesgos técnicos, científicos, jurídicos y organizativos

Los riesgos no se limitan al sesgo algorítmico. En sectores industriales regulados deben analizarse fallos de medición, selección, semántica, causalidad, seguridad, organización y responsabilidad.

12.1. Datos abundantes, pero no representativos

Los sistemas registran más normalidad que fallos. Un conjunto enorme puede contener pocos eventos críticos. Las métricas globales ocultarán rendimiento deficiente en clases raras. La solución combina muestreo dirigido, simulación, conocimiento mecanístico, validación prospectiva y vigilancia.

12.2. Atajos y variables proxy

El modelo puede aprender el instrumento, la planta, el fondo de una imagen o el patrón de documentación en lugar del fenómeno. Las pruebas deben cambiar condiciones no causales y evaluar explicaciones con expertos.

12.3. Confusión entre predicción y causalidad

Una variable predictiva no es necesariamente manipulable. Modificarla puede no cambiar el resultado. Las intervenciones requieren experimentación, conocimiento de proceso o métodos causales.

12.4. Deriva y obsolescencia

Cambios de proveedor, método, equipo, población o formulación alteran distribución. Se necesitan indicadores de entrada, salida y rendimiento, además de eventos de revalidación.

12.5. Automatización del sesgo histórico

Las decisiones pasadas pueden contener criterios inconsistentes. El modelo los reproduce con apariencia objetiva. Deben revisarse etiquetas y resultados, no solo variables.

12.6. Fuga de conocimiento y extracción

Modelos y servicios pueden memorizar, inferir o exponer secretos y datos personales. La protección incluye minimización, controles, pruebas de privacidad, segmentación y contratos.

12.7. Ataques a la integridad

Envenenamiento de datos, manipulación de sensores, ejemplos adversarios o cambios en pipelines pueden alterar decisiones. El modelo de amenazas debe abarcar datos, software, personas y cadena de suministro.

12.8. Deuda de datos

Duplicidades, códigos incompatibles y metadatos ausentes se acumulan como pasivo. Cuando se intenta automatizar, el coste aparece de forma abrupta. La deuda debe medirse y amortizarse por prioridad.

12.9. Obediencia algorítmica

La existencia de un humano en el circuito no garantiza control. Debe conocer límites, recibir información útil y poder detener o escalar.

12.10. Validación circular

Usar los mismos datos para seleccionar y confirmar produce optimismo. La independencia puede lograrse con conjuntos bloqueados, separación temporal, centros externos o estudios prospectivos.

Riesgo	Señal temprana	Control prioritario
Deriva	Cambian distribuciones, calibración o tasa de alertas.	Monitorización y umbrales de revalidación.
Fuga de datos	Uso de herramientas no autorizadas o exportaciones atípicas.	DLP, segregación, contratos y formación.
Atajo	Rendimiento cae al cambiar planta/equipo/fondo.	Validación externa y pruebas de perturbación.
Automatización indebida	El usuario no revisa o no comprende salidas.	Diseño de interfaz, autoridad y auditoría de decisiones.
Deuda de datos	Tiempo creciente de limpieza y contradicciones.	Catálogo, MDM, reglas y responsables.
Sobreajuste	Gran diferencia entre validación interna y prospectiva.	Separación estricta y evaluación independiente.
Evento raro	Alta exactitud con sensibilidad crítica baja.	Métricas por clase, simulación y muestreo dirigido.

Tabla 10. Indicadores tempranos de riesgo en sistemas de IA industrial.

13. Medición del retorno y del valor creado

El retorno no debe medirse por terabytes, modelos o paneles. Debe vincularse con decisiones y resultados. El valor directo incluye reducción de costes, desperdicio, tiempo de ciclo, consumo, rechazo, inventario y experimentos. El indirecto incluye aprendizaje acumulativo, preparación regulatoria, continuidad, reputación, barreras de entrada y opciones estratégicas.

En seguridad, el beneficio se expresa a menudo como pérdida evitada. La ausencia de incidentes dificulta atribución. Puede estimarse mediante reducción esperada de riesgo, tiempo de respuesta, cobertura, severidad y escenarios. La precisión aparente no debe transformarse sin más en euros: es necesario modelar cómo la salida modifica una actuación.

En I+D, los indicadores relevantes son tiempo hasta decisión, tasa de aprendizaje por experimento, reducción del espacio de búsqueda, porcentaje de resultados reutilizables y calidad de la transferencia. En fabricación, desviaciones evitadas, rendimiento, capacidad, tiempo de liberación y energía. En regulación, tiempo de preparación, consultas, inconsistencias y reutilización de evidencia.

La evaluación debe incluir coste total: captura, limpieza, integración, almacenamiento, licencias, ciberseguridad, validación, formación, monitorización, dependencia de proveedor y retirada. Los proyectos que ignoran mantenimiento confunden una demostración con una capacidad industrial.

Categoría	Indicador recomendado	Interpretación
I+D	Tiempo hasta decisión y experimentos evitados.	Mide velocidad de aprendizaje, no solo volumen de actividad.
Calidad	Tiempo de investigación, recurrencia y desviaciones detectadas antes de impacto.	Mide prevención y capacidad causal.
Seguridad alimentaria	Tiempo de alerta, cobertura, falsos negativos críticos y coste de control.	Relaciona predicción con gestión de riesgo.
Fabricación	Rendimiento, OEE, scrap, consumo y tiempo de liberación.	Mide efecto productivo integral.
Regulatorio	Tiempo de ensamblaje de evidencia, consultas y reutilización.	Mide preparación y trazabilidad.
Datos	Porcentaje con propietario, calidad, linaje y uso.	Mide conversión de registros en productos gobernados.
Modelos	Deriva, incidentes, abstenciones y acciones humanas.	Mide comportamiento real del sistema.
Finanzas	Valor esperado neto ajustado por riesgo y coste total de propiedad.	Evita valorar IA por ahorro bruto teórico.

Tabla 11. Indicadores para evaluar valor y retorno.

14. Hoja de ruta de implantación

La transformación debe avanzar por capacidades y casos de uso. El primer paso no es entrenar un modelo, sino identificar decisiones repetitivas, costosas o inciertas y determinar si existen datos y autoridad para actuar.

1. Inventariar decisiones críticas, fuentes de datos, derechos, sistemas y responsables. El inventario debe priorizar, no aspirar a catalogar todo antes de actuar.
2. Seleccionar casos de uso con valor mensurable, datos accesibles, riesgo controlable y propietario operativo. Deben incluir una línea base sin IA.
3. Construir el producto de datos: identidad, metadatos, calidad, semántica, linaje, seguridad, acceso y retención.
4. Desarrollar la solución más simple capaz de resolver el problema. Los modelos complejos deben justificar su ventaja sobre reglas, estadística o modelos mecánicos.
5. Validar de acuerdo con el contexto de uso, la consecuencia y los grupos relevantes. Incorporar incertidumbre y posibilidad de abstención.
6. Integrar en el flujo de trabajo con procedimientos, formación, revisión, registros, contingencia y autoridad clara.
7. Vigilar rendimiento, deriva, incidentes y beneficios. Cambios de datos, proceso, proveedor o finalidad deben activar evaluación.
8. Escalar mediante reutilización de componentes, estándares y productos de datos; no mediante clonación de pilotos sin gobierno.

14.1. Modelo de madurez

Nivel	Características	Riesgo dominante	Prioridad
0 · Fragmentado	Datos en documentos, hojas y sistemas aislados.	No se conoce la realidad informacional.	Inventario y control básico.
1 · Digitalizado	Captura electrónica, pero semántica y calidad locales.	Digitalizar errores y duplicidades.	Identidad, metadatos y responsables.
2 · Gobernado	Productos de datos, calidad, acceso y linaje.	Burocracia sin casos de uso.	Conectar gobierno con decisiones.
3 · Analítico	Modelos en producción para casos delimitados.	Pilotos sin mantenimiento o dependencia de proveedor.	Ciclo de vida y monitorización.
4 · Integrado	IA conectada con procesos, calidad y evidencia.	Riesgo sistémico y propagación de errores.	Resiliencia y validación cruzada.
5 · Adaptativo	Aprendizaje continuo controlado y experimentación cerrada.	Cambio no controlado y opacidad.	Gobierno de cambios y supervisión avanzada.

Tabla 12. Modelo de madurez de datos e IA en industrias reguladas.

14.2. Prioridades para los primeros doce meses

- Aprobar una política de datos e IA con clasificación de información y herramientas autorizadas.
- Crear un registro de sistemas y casos de uso, incluyendo soluciones adquiridas y funciones de IA embebidas.
- Designar propietarios de los productos de datos prioritarios.
- Seleccionar dos o tres casos con línea base, valor y riesgo medibles.
- Establecer plantillas de contexto de uso, ficha de datos, informe de validación y vigilancia.
- Revisar contratos de equipos, nube y software respecto a acceso, exportación, uso secundario y continuidad.
- Implantar formación diferenciada para dirección, usuarios, desarrolladores, calidad, jurídico y regulación.
- Crear un comité ligero de revisión con capacidad real de decisión y escalado.

15. Discusión: de la ventaja algorítmica a la ventaja epistemológica

La primera etapa de adopción empresarial premió a quienes podían adquirir capacidad computacional y talento especializado. La difusión de herramientas está reduciendo esa ventaja. El acceso a modelos no garantiza diferenciación cuando competidores utilizan servicios equivalentes.

La ventaja futura será epistemológica y organizativa: saber qué medir, cómo medirlo, qué representa, qué no representa, qué incertidumbre importa y qué experimento permite aprender. Esta capacidad se construye durante años y se incorpora en datos, procedimientos, vocabularios y juicio experto.

En la industria alimentaria, la ventaja consistirá en conectar variabilidad de materias primas, proceso, seguridad, calidad y mercado. En farmacia, en distinguir hipótesis, evidencia exploratoria y evidencia regulatoria, manteniendo integridad durante todo el ciclo. En química, en combinar datos, mecanismos y dominios de aplicabilidad. En Biotech, en cerrar el ciclo entre computación, laboratorio y biocuración.

La apertura y la protección no son objetivos incompatibles. El valor puede aumentar mediante intercambio, siempre que se preserve finalidad, atribución, seguridad y reparto de beneficios. El cierre absoluto reduce posibilidades de aprendizaje; la apertura indiscriminada destruye secretos, privacidad y posición competitiva. La gobernanza busca un régimen de acceso proporcional.

Existe también una dimensión ética y social. Los datos de pacientes, consumidores, trabajadores y agricultores no son meros insumos. Los sistemas pueden redistribuir poder, cargas y oportunidades. La legitimidad depende de finalidad, proporcionalidad, transparencia, derechos y posibilidad de impugnar consecuencias.

El concepto de explicabilidad debe tratarse con precisión. No siempre es posible explicar cada parámetro de un modelo complejo, ni una explicación local garantiza causalidad. Lo exigible es una transparencia funcional suficiente: propósito, datos, rendimiento, límites, incertidumbre, variables sensibles, supervisión y comportamiento ante casos no conocidos.

La adopción responsable no debe convertirse en inmovilismo. Exigir certeza absoluta impediría innovar. El enfoque correcto es proporcional al riesgo: controles más intensos cuando el sistema puede afectar seguridad, calidad, derechos o evidencia; experimentación más ágil en usos exploratorios con barreras claras para evitar que sus resultados se conviertan indebidamente en decisiones críticas.

16. Recomendaciones estratégicas

9. Tratar los datos como activos científicos gobernados, no como subproductos informáticos.
10. Definir el contexto de uso antes de seleccionar datos y modelo.
11. Priorizar procedencia, semántica y representatividad frente a volumen.
12. Registrar resultados negativos, excepciones, repeticiones y decisiones humanas.
13. Conservar datos brutos, metadatos, transformaciones, versiones y evidencia de validación.
14. Separar exploración, apoyo a decisión y uso regulatorio mediante barreras explícitas.
15. Integrar IA en los sistemas de calidad existentes; no crear un circuito paralelo.
16. Incluir derechos de acceso, portabilidad, confidencialidad y continuidad en contratos tecnológicos.
17. Monitorizar deriva y beneficios reales, no únicamente métricas de desarrollo.
18. Mantener autoridad humana efectiva, con capacidad de abstención, escalado y retirada.
19. Adoptar arquitecturas de compartición controlada cuando la concentración de datos no sea necesaria.
20. Medir y reducir deuda de datos como parte del riesgo empresarial.

17. Conclusiones

La inteligencia artificial no transforma automáticamente datos en conocimiento. Amplifica la capacidad de detectar patrones, generar hipótesis y optimizar decisiones, pero también amplifica errores de medición, sesgos, omisiones y falsas certezas. Su rendimiento industrial depende de una infraestructura epistemológica y organizativa previa.

En las industrias alimentaria, químico-farmacéutica y biotecnológica, el valor del dato se construye mediante calidad, contexto, procedencia, interoperabilidad, licitud, singularidad y accionabilidad. El volumen es una medida secundaria. Un conjunto pequeño, longitudinal y vinculado a resultados puede ser más valioso que millones de registros descontextualizados.

Los resultados negativos y las desviaciones forman parte del capital científico. Documentarlos permite aprender límites y evita repetir errores. La empresa que conserva de manera estructurada lo que funcionó, lo que no funcionó y en qué condiciones desarrolla una memoria industrial difícil de replicar.

La regulación europea no debe interpretarse como un conjunto de obstáculos aislados. *AI Act*, *Data Act*, *Data Governance Act*, RGPD, EHDS, secretos empresariales, derechos sobre bases y normas sectoriales configuran las condiciones bajo las cuales el valor puede realizarse legítimamente. La gobernanza jurídica es parte del diseño técnico.

La estrategia correcta comienza por una decisión, no por una herramienta. Pregunta qué incertidumbre debe reducirse, qué evidencia se necesita, qué datos la representan, qué consecuencias tendría el error y cómo se demostrará la validez del sistema.

El activo estratégico definitivo no es el dato bruto ni el algoritmo. Es la capacidad institucional para producir evidencia reproducible, tomar decisiones defendibles y aprender de forma acumulativa. En sectores donde la seguridad, la eficacia, la calidad y la innovación dependen de la prueba, gobernar los datos equivale a gobernar la capacidad de competir.

Conclusión ejecutiva

La empresa que compre IA sin construir gobierno del dato obtendrá automatizaciones frágiles. La empresa que convierta su experiencia en productos de datos trazables construirá una ventaja que los modelos comerciales, por sí solos, no pueden copiar.

Declaraciones editoriales

Autoría. José M. López es responsable de la concepción, análisis y redacción del manuscrito.

Afiliación. INTABIOTECH S.L., España.

Financiación. No se declara financiación externa específica para la elaboración de este artículo.

Conflictos de interés. El autor desarrolla actividad profesional en el ámbito de ingredientes, biotecnología y tecnología alimentaria. Esta vinculación debe ser considerada al valorar las implicaciones empresariales del trabajo. No se declara conflicto financiero relacionado con proveedores de plataformas de inteligencia artificial citados.

Disponibilidad de datos. El artículo es una revisión conceptual y normativa y no genera un conjunto de datos de investigación primario.

Uso de inteligencia artificial. Se emplearon herramientas digitales de apoyo a la búsqueda bajo revisión humana del contenido. La responsabilidad sobre afirmaciones, selección de fuentes y conclusiones corresponde al autor.

Nota jurídica. El análisis normativo tiene finalidad académica y estratégica. La clasificación y cumplimiento de un sistema concreto requieren evaluación individual del producto, uso, territorio, roles contractuales y normativa aplicable.

Glosario esencial

Término	Definición operativa
ALCOA+	Principios de integridad: atribuible, legible, contemporáneo, original, exacto, completo, consistente, perdurable y disponible.
Contexto de uso	Finalidad específica, decisión, población, entorno y consecuencia para los que se evalúa un modelo.
Data drift	Cambio en la distribución de los datos de entrada respecto al desarrollo.
Concept drift	Cambio en la relación entre variables y resultado.
Dominio de aplicabilidad	Región del espacio de datos o sustancias en la que una predicción es considerada válida.
FAIR	Principios para que datos y metadatos sean hallables, accesibles, interoperables y reutilizables.
Linaje	Registro de origen, movimientos, transformaciones y versiones de un dato.
Producto de datos	Conjunto gobernado y mantenido para usuarios y finalidades definidas.
RWD/RWE	Datos del mundo real/evidencia derivada de su análisis.

Término	Definición operativa
Seudonimización	Tratamiento que separa identificadores, sin excluir necesariamente la posibilidad de re-identificación.
Validación	Demostración documentada de aptitud para un uso previsto.

Referencias

- Comité Europeo de Protección de Datos. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*.
- Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C. (2022). *The value of data in digital-based business models: Measurement and economic policy implications* (OECD Economics Department Working Papers No. 1723). OECD Publishing. <https://doi.org/10.1787/d960a10c-en>
- Dessimoz, C., & Thomas, P. D. (2024). *AI and the democratization of knowledge*. *Scientific Data*, 11, 268. <https://doi.org/10.1038/s41597-024-03099-1>
- Durrant, A., Markovic, M., Matthews, D., May, D., Enright, J., & Leontidis, G. (2022). *The role of cross-silo federated learning in facilitating data sharing in the agri-food sector*. *Computers and Electronics in Agriculture*, 193, 106648. <https://doi.org/10.1016/j.compag.2021.106648>
- European Chemicals Agency. (2017). *Read-Across Assessment Framework (RAAF)*. European Union.
- European Chemicals Agency. (2026). *Data sharing under REACH*. European Union.
- European Chemicals Agency. (2026). *QSAR models: How to avoid unnecessary testing on animals*. European Union.
- European Food Safety Authority. (2025). *AI@EFSA*. EFSA.
- European Food Safety Authority. (2025). *Symposium on data readiness for artificial intelligence*. EFSA Supporting Publications, 22, EN-9434.
- European Food Safety Authority. (2026). *Data standardisation: Standard Sample Description and FoodEx2*. EFSA.
- European Medicines Agency. (2023). *Data Quality Framework for EU medicines regulation*. EMA/HMA.
- European Medicines Agency. (2023). *Guideline on computerised systems and electronic data in clinical trials* (EMA/INS/GCP/112288/2023).
- European Medicines Agency. (2024). *Reflection paper on the use of artificial intelligence in the medicinal product lifecycle* (EMA/CHMP/CVMP/83833/2023).
- European Medicines Agency. (2026). *Data Quality Framework for EU medicines regulation: Application to real-world data*.
- European Medicines Agency, & U.S. Food and Drug Administration. (2026). *Guiding principles of good AI practice in drug development*.
- Gavai, A., Bouzembrak, Y., Mu, W., van den Bulk, L. M., van Andel, A., van der Fels-Klerx, H. J., & Marvin, H. J. P. (2023). *Applying federated learning to combat food fraud in food supply chains*. *npj Science of Food*, 7, 46. <https://doi.org/10.1038/s41538-023-00220-3>
- International Council for Harmonisation. (2023). *ICH Q9(R1): Quality risk management*.
- International Council for Harmonisation. (2025). *ICH E6(R3): Guideline for good clinical practice*.
- International Organization for Standardization. (2023). *ISO/IEC 23894:2023. Information technology—Artificial intelligence—Guidance on risk management*.
- International Organization for Standardization. (2023). *ISO/IEC 42001:2023. Information technology—Artificial intelligence—Management system*.
- International Organization for Standardization. (2024). *ISO/IEC 5259 series. Artificial intelligence—Data quality for analytics and machine learning*.
- Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2022). *IoT, big data, and artificial intelligence in agriculture and food industry*. *IEEE Internet of Things Journal*, 9(9), 6305–6324. <https://doi.org/10.1109/JIOT.2020.2998584>
- Motzer, C., El-Moghazy, A., Allende, A., Gil, M. I., Weesepeel, Y., van Overbeek, L., van de Zedde, R., Bouzembrak, Y., Nitin, N., Ivanek, R., & Wiedmann, M. (2024). *Food safety related data analytics, digital, and artificial intelligence needs and opportunities in controlled environment agriculture*. *Food Protection Trends*, 44(6), 400–408. <https://doi.org/10.4315/FPT-24-017>
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- National Institute of Standards and Technology. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1).

- Notin, P., Rollins, N., Gal, Y., Sander, C., & Marks, D. S. (2024). Machine learning for functional protein design. *Nature Biotechnology*, 42, 216–228. <https://doi.org/10.1038/s41587-024-02127-0>
- Organisation for Economic Co-operation and Development. (2015). *Data-driven innovation: Big data for growth and well-being*. OECD Publishing. <https://doi.org/10.1787/9789264229358-en>
- Organisation for Economic Co-operation and Development. (2014). *Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models*. OECD Series on Testing and Assessment.
- Organisation for Economic Co-operation and Development. (2024). *(Q)SAR Assessment Framework: Guidance for the regulatory assessment of (quantitative) structure–activity relationship models and predictions (2nd ed.)*. OECD Publishing.
- Parlamento Europeo y Consejo. (1996). Directiva 96/9/CE, sobre la protección jurídica de las bases de datos. *Diario Oficial de las Comunidades Europeas*.
- Parlamento Europeo y Consejo. (2002). Reglamento (CE) n.º 178/2002, por el que se establecen los principios y los requisitos generales de la legislación alimentaria. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2006). Reglamento (CE) n.º 1907/2006, relativo al registro, la evaluación, la autorización y la restricción de las sustancias y mezclas químicas (REACH). *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2016). Directiva (UE) 2016/943, relativa a la protección de los conocimientos técnicos y la información empresarial no divulgados. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2016). Reglamento (UE) 2016/679, Reglamento General de Protección de Datos. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2022). Reglamento (UE) 2022/868, relativo a la gobernanza europea de datos. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2023). Reglamento (UE) 2023/2854, sobre normas armonizadas para un acceso justo a los datos y su utilización. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2024). Reglamento (UE) 2024/1689, por el que se establecen normas armonizadas en materia de inteligencia artificial. *Diario Oficial de la Unión Europea*.
- Parlamento Europeo y Consejo. (2025). Reglamento (UE) 2025/327, relativo al Espacio Europeo de Datos de Salud. *Diario Oficial de la Unión Europea*.
- Schwabe, D., Becker, K., Seyferth, M., Klaß, A., & Schaeffter, T. (2024). The METRIC-framework for assessing data quality for trustworthy AI in medicine: A systematic review. *npj Digital Medicine*, 7, 203. <https://doi.org/10.1038/s41746-024-01196-4>
- U.S. Food and Drug Administration. (2025). *Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: Draft guidance for industry and other interested parties*.
- van Breugel, B., Liu, T., Oglic, D., & van der Schaar, M. (2024). Synthetic data in biomedicine via generative artificial intelligence. *Nature Reviews Bioengineering*, 2, 991–1004. <https://doi.org/10.1038/s44222-024-00245-7>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- World Health Organization. (2021). *Guideline on data integrity*. WHO Technical Report Series No. 1033, Annex 4.
- Yang, H., Jiao, W., Zouyi, L., Diao, H., & Xia, S. (2025). Artificial intelligence in the food industry: Innovations and applications. *Discover Artificial Intelligence*, 5, 60. <https://doi.org/10.1007/s44163-025-00296-8>