

VEYXARA KNOWLEDGE SERIES

El ciclo de los datos (El verdadero combustible de la IA)

Una guía práctica para comprender sus fundamentos y aplicaciones

Autor: Alfredo Arévalo
Publicación: Veyxara Core

Plataforma oficial: www.veyxara.com

Repositorio de conocimiento:
https://github.com/alfredoarevalo69/recursos-publicos

Edición: 2026

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados



El ciclo de los datos: el verdadero combustible de la IA

Cuando comenzamos a utilizar herramientas de Inteligencia Artificial, es fácil tener la impresión de que estamos interactuando con una entidad que lo sabe prácticamente todo, como si tuviera una mente propia capaz de comprender el mundo de la misma manera que un ser humano.

Pero detrás de esa capacidad existe algo mucho más interesante: grandes cantidades de información, modelos matemáticos y procesos diseñados para identificar patrones, aprender de los datos y generar resultados.

La calidad de una solución de IA depende de múltiples factores: los datos utilizados, el diseño y la arquitectura del modelo, el proceso de entrenamiento, la capacidad de procesamiento, las instrucciones y el contexto disponibles, y la forma en que se evalúan sus resultados. Los datos son uno de sus componentes fundamentales, pero no el único.

En el mundo tecnológico existe una expresión muy conocida: **GIGO (Garbage In, Garbage Out)**, que significa «entra basura, sale basura». En otras palabras, si alimentamos un sistema con información incorrecta, incompleta o de mala calidad, podemos obtener resultados poco confiables.

En esta guía vamos a conocer, sin tecnicismos complejos, cómo se preparan, procesan y utilizan los datos para hacer posible el funcionamiento de la Inteligencia Artificial.

1. ¿Qué es realmente un Dataset (conjunto de datos)?

Cuando escuchamos la palabra *dataset*, podemos imaginar grandes bases de datos o estructuras complejas. Sin embargo, el concepto es mucho más sencillo.

Un dataset es un conjunto de datos recopilados y organizados para un propósito determinado. Puede contener información numérica, textos, imágenes, grabaciones de audio, videos y muchos otros formatos.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

Dependiendo de cómo estén organizados, podemos encontrar tres grandes categorías:

Ejemplo sencillo: una tabla con tres columnas —producto, precio y cantidad— puede formar un pequeño dataset de ventas. Si añadimos muchas filas con registros de distintas ventas, tendremos un conjunto de datos que puede analizarse para identificar productos más vendidos, tendencias o valores atípicos.

- **Datos estructurados:** son aquellos que tienen una organización definida, generalmente en filas y columnas. Por ejemplo, una hoja de Excel con nombres, fechas, números de identificación, productos y precios.
- **Datos semiestructurados:** contienen cierta organización mediante etiquetas o estructuras, pero no siguen necesariamente el formato rígido de una tabla tradicional. Algunos ejemplos son los archivos JSON, XML y determinados registros de sistemas.
- **Datos no estructurados:** incluyen información que no tiene una organización tabular predefinida, como documentos PDF, correos electrónicos, fotografías, grabaciones de voz y videos.

Para que un modelo de IA pueda trabajar con estos diferentes formatos, la información debe pasar por procesos que permitan representarla y convertirla en estructuras que el sistema pueda procesar.

Por ejemplo, en el procesamiento del lenguaje natural se utiliza la tokenización: un proceso que divide el texto en unidades más pequeñas llamadas tokens. Según el modelo y el idioma, un token puede representar una palabra, parte de una palabra o un signo. Después, esas unidades se representan mediante valores numéricos para que el modelo pueda procesarlas.

2. El trabajo invisible: limpieza y preparación de datos

Uno de los errores más comunes al hablar de Inteligencia Artificial es pensar que todo consiste en desarrollar algoritmos sofisticados y ponerlos a funcionar.

En realidad, antes de utilizar los datos es necesario realizar un trabajo importante de preparación para garantizar que la información tenga una calidad adecuada.

Imaginemos que una empresa quiere implementar un sistema de IA para reconocer y clasificar facturas automáticamente.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

¿Qué sucedería si le proporcionamos documentos duplicados, fotografías borrosas, fechas con formatos diferentes o información incompleta?

El sistema podría interpretar incorrectamente los documentos, generar clasificaciones equivocadas o entregar resultados poco confiables.

Por eso, la preparación de datos puede incluir actividades como:

- **Eliminar duplicados:** identificar registros repetidos que pueden alterar los resultados.
- **Tratar información faltante:** identificar valores incompletos y determinar cómo gestionarlos.
- **Corregir inconsistencias:** detectar errores y diferencias en los registros.
- **Estandarizar formatos:** unificar fechas, monedas, nombres y otras estructuras.
- **Validar la información:** comprobar que los datos sean coherentes y adecuados para el propósito del sistema.

Este proceso no es simplemente una tarea administrativa. Es una parte fundamental para conseguir que los modelos de IA produzcan resultados útiles y confiables.

3. Entrenamiento y sesgos (Bias): cuando la IA aprende de nuestros datos

Una vez que los datos han sido preparados, podemos utilizarlos en diferentes procesos, entre ellos el entrenamiento de modelos de Inteligencia Artificial.

Durante el entrenamiento, un modelo analiza grandes cantidades de ejemplos y ajusta sus parámetros para identificar patrones y relaciones que le permitan realizar determinadas tareas.

Sin embargo, aquí aparece un aspecto importante: los datos utilizados pueden contener errores, desequilibrios o sesgos presentes en la información original.

A esto lo conocemos como **sesgo (Bias)**.

Imaginemos una empresa que desarrolla un sistema para apoyar la selección de candidatos a puestos directivos de tecnología.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

Si los datos históricos utilizados para entrenar el modelo muestran que la mayoría de las personas seleccionadas para esos cargos eran hombres, el sistema podría aprender patrones que favorezcan ese comportamiento histórico.

Como consecuencia, podría perjudicar injustamente a candidatas que cuentan con las mismas capacidades y experiencia.

Esto no significa que todos los modelos necesariamente reproduzcan esos resultados. El comportamiento también depende del diseño del sistema, los datos seleccionados, los criterios de evaluación y los controles implementados.

Por eso, la diversidad y representatividad de los datos son importantes tanto desde el punto de vista técnico como ético.

4. No todos los datos que recibe una IA se utilizan para entrenarla

Este es un concepto que vale la pena comprender, especialmente cuando comenzamos a utilizar herramientas de IA en nuestro trabajo diario.

Existe una diferencia entre la información utilizada para entrenar un modelo y aquella que recibe posteriormente para responder una consulta.

Podemos distinguir tres conceptos:

- **Datos de entrenamiento:** son los ejemplos utilizados durante el desarrollo del modelo para que aprenda patrones y ajuste sus parámetros.
- **Datos de contexto:** son las instrucciones y la información que se proporcionan al modelo durante una interacción para orientar su respuesta.
- **Datos de recuperación (RAG):** corresponden a información externa que un sistema consulta cuando necesita complementar sus respuestas con documentos o fuentes específicas.

Por ejemplo, imaginemos que una empresa tiene cientos de documentos técnicos sobre sus servidores, procedimientos y políticas internas.

Mediante una arquitectura RAG, podría implementar un asistente de IA que consulte esos documentos para responder preguntas de sus colaboradores.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

En términos sencillos, el flujo puede entenderse así:

1. El usuario realiza una pregunta al asistente.

2. El sistema busca en las fuentes autorizadas los fragmentos de información relacionados.

3. Recupera los fragmentos más relevantes y los incorpora como contexto.

4. El modelo utiliza la pregunta y ese contexto para elaborar una respuesta.

5. Cuando es posible, el sistema presenta referencias para que la persona pueda verificar la información.

Importante: RAG no garantiza por sí solo que la respuesta sea correcta. La calidad depende, entre otros factores, de los documentos disponibles, la recuperación de información, las instrucciones y la evaluación del resultado.

Esto permite aprovechar información empresarial sin que necesariamente sea necesario volver a entrenar el modelo con todos esos documentos.

Es una de las aplicaciones más interesantes de la IA en entornos corporativos.

5. Evaluación y retroalimentación: un proceso continuo

El trabajo con los datos y los modelos de IA no termina cuando finaliza el entrenamiento.

Antes de utilizar un modelo en un entorno real, es necesario evaluar su comportamiento para determinar si cumple con los objetivos para los que fue desarrollado.

Esta evaluación puede incluir:

Una respuesta bien redactada puede parecer convincente y aun así contener errores. Por eso, no basta con evaluar cómo suena una respuesta: también hay que contrastar su contenido con información confiable y comprobar si cumple el objetivo esperado.

- **Precisión de los resultados:** comprobar si las respuestas o predicciones son correctas.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

- **Detección de errores:** identificar situaciones en las que el modelo no responde adecuadamente.
- **Evaluación de sesgos:** comprobar si existen resultados desproporcionados o injustos.
- **Seguimiento del comportamiento:** observar si su rendimiento se mantiene cuando cambian los datos o las condiciones de uso.

Con el tiempo, las necesidades pueden cambiar, aparecen nuevos datos y pueden identificarse errores que no se habían detectado inicialmente.

Por eso, los sistemas de IA requieren seguimiento y, cuando corresponde, ajustes o nuevas etapas de entrenamiento y evaluación.

La Inteligencia Artificial no debe entenderse como una solución que se configura una sola vez y funciona indefinidamente sin supervisión.

6. Seguridad y privacidad: proteger los datos también es parte de la IA

Hay un aspecto que no podemos pasar por alto: los datos utilizados por una solución de Inteligencia Artificial pueden contener información sensible o confidencial.

En una empresa, por ejemplo, los documentos pueden incluir información de clientes, empleados, contratos, configuraciones de infraestructura o procedimientos internos.

Pensemos en una situación cotidiana: una persona recibe un archivo con datos de clientes y lo carga en una herramienta pública de IA para pedir un resumen. Aunque la intención sea ahorrar tiempo, podría estar compartiendo información personal o confidencial con un servicio externo.

Antes de utilizar una herramienta de IA, conviene revisar las políticas de la organización y las condiciones del servicio. Si no existe autorización, evita introducir datos personales, credenciales, información de clientes o documentos internos. Cuando sea posible, utiliza datos ficticios, anonimizados o minimizados.

Por eso, antes de implementar cualquier solución de IA es importante considerar:

- **Privacidad:** evitar la exposición innecesaria de datos personales.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

- **Confidencialidad:** proteger la información empresarial y restringir su uso según corresponda.
- **Control de acceso:** garantizar que únicamente las personas y sistemas autorizados puedan consultar la información.
- **Trazabilidad:** mantener mecanismos que permitan conocer el origen y los cambios de los datos.
- **Gobierno de datos:** establecer políticas y responsabilidades sobre cómo se recopila, almacena, utiliza y protege la información.

Una solución de IA puede ser técnicamente avanzada, pero si no protege adecuadamente la información que utiliza, puede convertirse en un riesgo para la organización.

7. Automatización y herramientas "No-Code" (IA en tu día a día)

Cuando se habla de automatizar procesos con Inteligencia Artificial, la primera imagen que a muchos se le viene a la mente es la de un programador escribiendo cientos de líneas de código en pantallas negras. Te confieso que durante un tiempo yo también pensé que para poner a la IA a trabajar "de fondo" en mis tareas diarias necesitaba dominar lenguajes de programación complejos.

La tecnología ha reducido considerablemente esa barrera de entrada. Hoy existen herramientas No-Code (sin código) y Low-Code (con poco código) que permiten construir flujos de trabajo mediante interfaces visuales. Sin embargo, no significa que automatizar sea simplemente conectar bloques: también es necesario comprender el proceso, los permisos, la seguridad, las integraciones y el manejo de errores.

En esta guía te voy a mostrar cómo pasar de chatear manualmente con la IA a delegarle tareas para que las ejecute de forma automática.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

El concepto No-Code: Trabajar con bloques, no con código

El enfoque No-Code se basa en una idea sencilla: representar un proceso mediante pasos visuales, sin tener que escribir código para cada operación. Conocer el proceso que se quiere mejorar sigue siendo indispensable; la herramienta facilita su construcción, pero no sustituye el análisis.

En lugar de programar desde cero, utilizas interfaces visuales donde arrastras y conectas aplicaciones que ya usas a diario (tu correo, tus hojas de cálculo, tu gestor de tareas) con modelos de IA.

La IA deja de ser una pestaña abierta en tu navegador a la que le haces preguntas de vez en cuando, y se convierte en un **asistente en segundo plano** que procesa información en el momento exacto en que ocurre un evento.

Casos de uso prácticos que puedes aplicar hoy

No necesitas inventar sistemas complejos para ver el impacto. Aquí tienes tres ejemplos sencillos que transforman el trabajo operativo:

- **Resumen y extracción de documentos:** En lugar de leer reportes de 40 páginas o actas de reuniones eternas, puedes configurar un flujo donde subes un PDF y la IA extrae automáticamente los puntos clave, las decisiones tomadas y las tareas pendientes, entregándote un resumen estructurado en minutos.
- **Análisis inteligente en hojas de cálculo:** Conectar la IA a herramientas como Google Sheets o Excel te permite clasificar cientos de comentarios de clientes (en positivos, neutros o negativos), estandarizar formatos de datos o detectar anomalías sin necesidad de escribir fórmulas complejas.
- **Generación de respuestas y borradores en lote:** Si recibes correos con solicitudes recurrentes, la IA puede leer el mensaje entrante, redactar un borrador de respuesta adaptado al contexto del cliente y dejártelo listo para que tú solo tengas que revisarlo y darle a "enviar".

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados

8. La lógica de un flujo automatizado (Disparador + Acción)

Para empezar a pensar en modo automatización, puedes estructurar una tarea con una regla lógica sencilla: «Cuando pase X, realiza Y y entrega Z». En un flujo real también conviene definir qué hacer si falta información, ocurre un error o el resultado necesita revisión humana.

1. **El disparador (Trigger):** Es el evento que inicia todo. Por ejemplo: *"Llega un nuevo correo electrónico con un archivo adjunto"*.
2. La acción o procesamiento: es el trabajo que realiza el flujo. Por ejemplo: «La IA analiza el archivo adjunto y extrae el monto total de la factura y el nombre del proveedor».
3. **El resultado final:** Dónde se guarda o envía el trabajo. Por ejemplo: *"Guarda esos dos datos en una fila de mi hoja de cálculo de gastos"*.

Al conectar estos tres pasos, acabas de eliminar un proceso manual que antes te tomaba 5 o 10 minutos cada vez que ocurría.

La regla de oro: El peligro de automatizar a ciegas

Automatizar es fantástico, pero delegar el 100% del control sin supervisión es un riesgo innecesario.

Como vimos en las guías anteriores, los modelos de lenguaje pueden cometer errores o interpretar mal una instrucción si el contexto es ambiguo. Por eso, mi recomendación siempre es mantener el **factor humano en el ciclo** (*Human-in-the-loop*).

Utiliza la automatización para reducir el trabajo repetitivo y facilitar las tareas operativas: extraer información, organizar datos o preparar borradores. Mantén la revisión humana en las actividades que requieren criterio, validación o toma de decisiones, especialmente cuando pueden afectar a personas, clientes, finanzas o servicios importantes.

Una buena automatización no es la que elimina todos los controles, sino la que hace el proceso más eficiente sin perder trazabilidad, seguridad ni capacidad de corregir errores.

© 2026 Veyxara Core · www.veyxara.com · Todos los derechos reservados



Conclusión: la IA comienza mucho antes de generar una respuesta

Cuando utilizamos una herramienta de Inteligencia Artificial, normalmente solo vemos el resultado final: una respuesta, una imagen, un análisis o una recomendación.

Pero detrás de ese resultado existe un conjunto de procesos que comienzan con los datos, continúan con su preparación y procesamiento, y requieren evaluación, supervisión y controles permanentes.

Comprender este proceso nos permite dejar de ver la IA únicamente como una herramienta para hacer preguntas y empezar a entenderla como una tecnología que depende de información, modelos y procesos bien diseñados.

La Inteligencia Artificial puede ofrecer capacidades muy valiosas, pero su utilidad depende de la combinación de datos adecuados, modelos bien diseñados, procesos evaluados y una gestión responsable de la información.

VEYXARA CORE
Conocimiento tecnológico, infraestructura y evolución digital.

https://veyxara.com/

Repositorio público de conocimiento:
https://github.com/alfredoarevalo69/recursos-publicos

© 2026 Veyxara Core. Todos los derechos reservados.