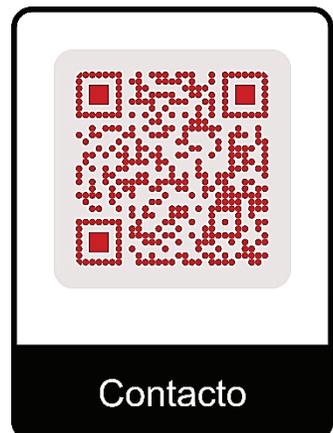

ACERCA DE LA AUTORA

Beatriz Coronado García

Máster en Prevención de Riesgos Laborales (3 especialidades) por la Universidad Francisco de Vitoria (2020-2021). Intensivo de experto en desarrollo de aplicaciones web por la Universidad San Jorge–SEAS (2021-2022). Grado en Sociología por la Universidad Rey Juan Carlos (2013-2017).

Profesional autónoma especializada en la gestión de proyectos editoriales y desarrollo de contenido formativo, con experiencia en tecnologías educativas y desarrollo web. Actualmente, trabaja con varias editoriales. Tiene experiencia en la utilización de diversas IA en el entorno laboral: ChatGPT 4.0, Copilot, Perplexity, Gemini y Midjourney, así como en el manejo de Microsoft 365 Business Standard. Además, cuenta con amplios conocimientos en lenguajes de programación como HTML5, CSS3 y JavaScript, y en sistemas de gestión de contenidos como WordPress.





INTRODUCCIÓN

En los últimos años, el término “Big Data” ha ganado protagonismo en prácticamente todos los sectores. No se trata solo de una moda tecnológica, sino de una transformación profunda en la forma de entender, procesar y aprovechar los datos. Este manual está pensado para ayudar a comprender, de forma clara y progresiva, qué es realmente el Big Data, cómo ha evolucionado desde el Business Intelligence tradicional y por qué está revolucionando tantos ámbitos.

<p> Capítulo 1 – Conceptos básicos y evolución</p> <ul style="list-style-type: none"> De BI a Big Data Qué es Big Data Hadoop Las "V" Real Time y Cloud	<p> Capítulo 2 – Arquitectura y tecnologías</p> <ul style="list-style-type: none"> HDFS y MapReduce Lenguajes clave ETL: Flume, Sqoop, Hive Kafka, HBase, Redis Spark y seguridad
<p> Capítulo 3 – Ciencia de Datos e IA</p> <ul style="list-style-type: none"> Introducción a IA Lenguajes: R y Python Algoritmos supervisados No supervisados Textos, imágenes y visualización	<p> Capítulo 4 – Aplicaciones y futuro</p> <ul style="list-style-type: none"> Open Data público Uso en empresas Data for Good Impacto y ética

A lo largo de los distintos capítulos, se abordan desde los conceptos más básicos hasta las aplicaciones más actuales, incluyendo las tecnologías más usadas, la relación con la inteligencia artificial y el impacto social que puede llegar a tener. Todo explicado con un lenguaje directo y con ejemplos cercanos, para que cualquier persona pueda seguirlo, independientemente de su formación previa.

1

ADQUISICIÓN Y DOMINIO DE CONCEPTOS BÁSICOS Y DE CONOCIMIENTOS SOBRE LOS AVANCES EN BIG DATA

Antes de profundizar en herramientas y tecnologías, es importante entender de dónde venimos y hacia dónde vamos en el mundo de los datos. Este primer capítulo sirve como base para familiarizarse con la evolución del tratamiento de la información: desde el Business Intelligence más tradicional hasta el uso del Big Data en contextos como la navegación web, la televisión o la geolocalización.



También se explican los conceptos clave que definen el Big Data —esas famosas “V”— y se presenta Hadoop, una tecnología que marcó un antes y un después en la forma de procesar datos en paralelo. Además, se dan unas primeras pinceladas de nuevos modelos de trabajo como el procesamiento en tiempo real o el uso de la nube.

1.1 CÓMO EVOLUCIONA EL BI TRADICIONAL AL BIG DATA (NAVEGACIÓN WEB, GEOLOCALIZACIÓN, AUDIENCIAS TV,...)

Antes de que se hablara tanto de Big Data, lo más habitual en las empresas era usar lo que se conoce como **Business Intelligence (BI) tradicional**. Esto consistía, básicamente, en recopilar datos estructurados (es decir, bien organizados en filas y columnas) que venían de bases de datos relacionales como SQL o incluso hojas de cálculo tipo Excel. A partir de ahí, se generaban **informes e indicadores** que ayudaban a analizar lo que había ocurrido: ventas del último mes, evolución del stock, rendimiento de un producto, etc. Era un enfoque bastante descriptivo, enfocado a entender el pasado y tomar decisiones sobre lo que ya había sucedido.

Este modelo funcionó muy bien durante años, pero tenía algunas **limitaciones importantes**. Por ejemplo, estaba pensado para **volúmenes de datos moderados**. Cuando las empresas empezaron a generar más información, el sistema se quedaba corto. Además, solo podía manejar datos bien estructurados, dejando fuera todo lo que no encajara en una tabla tradicional. El procesamiento era lento, ya que los informes se generaban en lotes (lo que se llama “batch”), muchas veces con días de retraso. Y por último, no permitía reaccionar con agilidad a cambios que ocurrieran en el momento.

Con el tiempo, y sobre todo con la llegada masiva de internet, los smartphones y los sensores conectados, los datos empezaron a crecer no solo en cantidad, sino también en **variedad y velocidad**. Ya no solo se trataba de ventas o inventario: ahora se recogían **clics en páginas web, ubicaciones GPS, interacciones en redes sociales, uso de apps móviles**, y mucho más. Apareció entonces la necesidad de un sistema que no solo recogiera más datos, sino que pudiera analizarlos **casi al instante** y de forma más inteligente, detectando patrones, haciendo predicciones y permitiendo decisiones más rápidas. Ahí es donde entra el Big Data.



Para ver esta evolución de forma clara, basta con pensar en algunos **ejemplos reales**. Cuando visitas una tienda online y te muestra productos que te podrían gustar, eso no se basa en informes antiguos, sino en el análisis de tu comportamiento en tiempo real. Lo mismo pasa con las **recomendaciones según tu ubicación**, como cuando una app te sugiere un restaurante cercano o te avisa de un atasco. También se ve en cómo las plataformas de streaming, como Netflix o Movistar+, ajustan sus contenidos en función de los gustos de su audiencia prácticamente al momento.

Este cambio de paradigma no ha sido solo una cuestión tecnológica, sino también una evolución en la **forma de pensar y gestionar la información** dentro de las organizaciones. Mientras que el BI tradicional requería tiempos largos de preparación, validación y generación de informes, el Big Data introduce un enfoque mucho más dinámico y automatizado. Ahora, los datos pueden analizarse de forma **continua y en streaming**, lo que permite detectar tendencias o anomalías al instante.

Además, una de las grandes diferencias es el tipo de datos que se pueden tratar. Antes se trabajaba casi exclusivamente con datos internos de la empresa: ventas, almacén, recursos humanos... Hoy en día, gracias al Big Data, es posible incorporar **fuentes externas y no estructuradas**, como comentarios en redes sociales, vídeos, audios, datos de sensores, clics en páginas web o incluso imágenes. Esto permite tener una visión mucho más rica y completa de lo que está ocurriendo, tanto dentro como fuera de la organización.



Otra ventaja importante es que el Big Data **no se limita a describir lo que ha pasado**, sino que permite generar modelos que **anticipen comportamientos futuros**. Por ejemplo, se pueden predecir patrones de consumo, detectar riesgos antes de que ocurran o personalizar servicios de forma muy precisa. Esto se traduce en decisiones más inteligentes, más rápidas y mucho mejor ajustadas a la realidad del momento.

Volviendo al ejemplo de la publicidad personalizada, ya no se lanza el mismo anuncio a todo el mundo: el sistema aprende del comportamiento del usuario y adapta los contenidos en función de sus intereses reales. O en el caso de la televisión y el streaming, las plataformas saben qué contenidos son más populares en cada franja horaria, región o grupo de edad, y pueden modificar su programación o sus recomendaciones en tiempo real.

Esta transición del BI clásico al enfoque Big Data representa una transformación profunda. No se trata solo de cambiar herramientas, sino de adoptar una nueva mentalidad basada en la agilidad, la diversidad de fuentes y el valor estratégico de los datos. Con el Big Data, los datos dejan de ser algo estático para convertirse en un recurso vivo, que evoluciona al ritmo de la actividad digital, y que permite a las organizaciones tomar decisiones más rápidas, más informadas y eficaces.

1.2 EL BIG DATA COMO SOLUCIÓN AL TRATAMIENTO MASIVO DE DATOS

Definición e historia desde su invención

Hablar de Big Data es hablar de una nueva forma de entender y trabajar con los datos. De forma sencilla, se puede decir que el Big Data es un **conjunto de tecnologías, métodos y enfoques** que permite procesar y analizar cantidades enormes de datos, que además son muy variados, se generan a gran velocidad y pueden aportar mucho valor si se gestionan bien. No estamos hablando solo de tener muchos datos, sino de **ser capaces de sacarles partido en tiempo y forma**.

Evolución del Big Data

● Años 90

La NASA y otras instituciones científicas comienzan a enfrentarse a grandes volúmenes de datos difíciles de procesar.

● Inicio de los 2000

Explosión de datos por internet, smartphones y redes sociales. Las herramientas tradicionales ya no son suficientes.

● 2004

Google presenta MapReduce y GFS, base del procesamiento distribuido moderno.

● 2006

Nace Hadoop, que democratiza el acceso al procesamiento masivo de datos gracias a su arquitectura abierta.

● 2010-2015

El ecosistema Big Data crece con herramientas como Spark, Hive, HBase y Kafka. Su uso se extiende a más sectores.

● Desde 2015

Big Data se integra en banca, salud, transporte, comercio... Aparecen nuevos perfiles profesionales como científicos de datos.

● Actualidad

Big Data convive con el BI tradicional y permite análisis predictivo, personalización y toma de decisiones en tiempo real.

El origen del problema que dio lugar al Big Data fue bastante claro: **los datos empezaron a crecer más rápido de lo que las herramientas tradicionales podían gestionar**. Al principio, las empresas trabajaban con datos estructurados y bases de datos relativamente pequeñas, pero con la llegada de internet, los smartphones y los sistemas conectados, esa cantidad se disparó. Empezamos a generar datos sin parar: correos electrónicos, registros web, videos, sensores, redes sociales, compras online... Y claro, las herramientas que funcionaban bien con unos pocos gigas o incluso teras, **empezaron a quedarse cortas**.

Aunque el término “Big Data” se puso de moda más adelante, **ya en los años 90** algunos sectores empezaron a enfrentarse a este problema. La **NASA, por ejemplo, y otras instituciones científicas** ya generaban cantidades de datos difíciles de manejar, especialmente en áreas como la meteorología o la exploración espacial. Sin embargo, fue en los **años 2000** cuando el concepto empezó a cobrar fuerza, sobre todo de la mano de empresas tecnológicas como **Google, Facebook, Amazon o Yahoo**, que se dieron cuenta de que necesitaban **nuevas formas de almacenar y procesar datos**, y que además necesitaban hacerlo muy rápido.



Las soluciones que surgieron para afrontar este reto fueron innovadoras y rompieron con los modelos tradicionales. Por ejemplo, se empezó a utilizar **almacenamiento distribuido**, es decir, guardar los datos no en un único servidor, sino en muchos equipos conectados entre sí. También se desarrollaron técnicas de

procesamiento paralelo, donde las tareas se dividen y se resuelven a la vez por varios nodos, lo que permite trabajar con grandes volúmenes en mucho menos tiempo. Y no solo se buscaba analizar lo que ya había pasado, sino también anticiparse: por eso, **el análisis predictivo** se convirtió en una pieza clave del enfoque Big Data.

Un hito muy importante en esta historia fue la publicación de **Google en 2004** sobre dos tecnologías propias:

- **MapReduce** (para procesar datos en paralelo).
- **Google File System (GFS)**, su sistema de archivos distribuido.

Este trabajo inspiró a muchos desarrolladores y dio lugar, poco después, al nacimiento de **Hadoop**, un proyecto de código abierto que permitió a otras empresas aplicar estos conceptos sin tener que reinventar la rueda.

Saber más...

¿Qué es MapReduce?

MapReduce es un modelo de programación desarrollado por Google para procesar grandes volúmenes de datos de forma paralela. Esto significa que divide el trabajo en pequeñas tareas, las reparte entre varios ordenadores (nodos) que trabajan al mismo tiempo, y luego junta los resultados.

Se compone de dos fases principales:

1. Map:

Toma los datos originales y los transforma en pares clave-valor. Por ejemplo, si tienes millones de comentarios de usuarios, esta fase podría contar cuántas veces aparece cada palabra.

2. Reduce:

Recoge todos esos pares clave-valor generados y los agrupa para sacar un resultado final. Siguiendo el ejemplo anterior, sumaría las veces que aparece cada palabra en todos los comentarios.

Este sistema permite procesar datos a gran escala de forma eficiente, sin que un solo servidor tenga que hacerlo todo. Es como repartir las tareas entre un grupo grande de personas y luego combinar sus respuestas para obtener la solución.

Hoy en día, aunque MapReduce sigue utilizándose, ha sido superado en muchos casos por tecnologías más rápidas y flexibles como Apache Spark, que también permite procesamiento distribuido, pero en memoria (más veloz), y con más opciones de análisis.

¿Qué es Google File System (GFS)?

Google File System (GFS) es un sistema de archivos distribuido creado también por Google, diseñado para almacenar enormes cantidades de datos en múltiples servidores. A diferencia de un sistema tradicional que guarda todos los archivos en un solo lugar, GFS divide los archivos en fragmentos (chunks) y los distribuye en varios servidores.

¿Para qué? Para lograr tres cosas:

➤ Escalabilidad:

Si necesitas más espacio, simplemente añades más servidores.

➤ Tolerancia a fallos:

Cada fragmento de archivo se guarda en varias copias, por si alguno falla.

➤ Rendimiento:

Puedes leer o escribir en varias partes del archivo al mismo tiempo.

GFS fue la base de inspiración para HDFS (Hadoop Distributed File System), que es el sistema de archivos usado por Hadoop. Hoy en día, GFS ha evolucionado internamente dentro de Google hacia Colossus, una versión más moderna y eficiente.

El término **Big Data se empezó a utilizar cada vez más**, primero en el mundo tecnológico, y poco después en sectores tan variados como la banca, la salud, el transporte o el marketing. No era solo una moda: era una respuesta real a una necesidad cada vez más común.

Desde ese momento, el Big Data ha ido evolucionando constantemente, tanto en capacidades técnicas como en su aplicación práctica. Lo que comenzó como una solución para almacenar y procesar grandes volúmenes de datos en entornos técnicos

o científicos, hoy se ha extendido a **casi todos los sectores de la sociedad**. Empresas de logística, hospitales, supermercados, administraciones públicas, plataformas de contenido o bancos, entre muchos otros, han adoptado el Big Data como parte de su día a día.



Lo interesante es que el crecimiento del Big Data no ha venido solo de la cantidad de datos generados, sino también de la **variedad de fuentes y formatos**. Ya no se trata solo de recoger números en tablas, sino de **entender opiniones, imágenes, vídeos, movimientos, comportamientos y contextos**, todo a la vez y en tiempo real. Gracias a esto, las organizaciones pueden ser mucho más precisas en lo que hacen: ajustan sus productos, detectan problemas antes de que se conviertan en crisis, mejoran sus servicios y ofrecen experiencias más personalizadas.

También, han aparecido **nuevas profesiones y perfiles técnicos** relacionados con esta revolución: ingenieros de datos, analistas, científicos de datos, arquitectos de datos... Todos ellos trabajan para que esa información masiva y a veces desordenada tenga sentido y se convierta en algo útil. Porque tener datos ya no es el reto; el reto es **saber qué hacer con ellos** y cómo traducirlos en decisiones.

En este proceso, el Big Data no ha sustituido las técnicas anteriores, como el BI tradicional que vimos antes, sino que **las complementa y las potencia**. Mientras el BI nos ayuda a entender qué ocurrió y por qué, el Big Data nos permite descubrir patrones ocultos, predecir comportamientos y actuar con agilidad. Por eso, muchas organizaciones combinan ambos enfoques según sus necesidades.

El Big Data no es una herramienta concreta, ni una única tecnología, sino una forma de trabajar con datos en entornos complejos, cambiantes y en constante crecimiento. Su historia muestra cómo la tecnología responde a los desafíos reales del mundo digital, y cómo, con creatividad y colaboración, se pueden encontrar soluciones que transforman la manera en que pensamos, decidimos y actuamos. Lo mejor es que esta historia no ha terminado: el Big Data sigue creciendo, abriendo nuevas posibilidades en campos como la inteligencia artificial, la automatización, la salud o el medio ambiente. Y apenas estamos viendo el comienzo.

1.3 HADOOP COMO REVOLUCIÓN PARA EL TRATAMIENTO PARALELO DE DATOS MASIVO

Como sabemos, durante muchos años, trabajar con grandes volúmenes de datos era una tarea complicada y costosa. Los sistemas tradicionales no estaban pensados para manejar tanta cantidad de información, y las soluciones que existían eran caras, lentas o poco flexibles. En este contexto, **Hadoop** supuso un antes y un después. Se trata de un **marco de trabajo open source** que permite almacenar y procesar datos masivos en lo que se conoce como **clústeres distribuidos**, es decir, redes de ordenadores que trabajan juntos como si fueran uno solo.



Lo que hizo especial a Hadoop fue su enfoque: en lugar de depender de servidores grandes y caros, como era lo habitual, proponía usar **equipos más sencillos y asequibles (commodity hardware)**, distribuyendo las tareas entre ellos. Así se conseguía un sistema mucho más económico y escalable. Si necesitabas más capacidad, no hacía falta cambiar toda la infraestructura: bastaba con **añadir más máquinas** al clúster.



Además, Hadoop introdujo el concepto de **procesamiento paralelo de forma sencilla**. Esto significa que cuando se recibe una gran cantidad de datos, el sistema los divide en partes y los procesa al mismo tiempo en diferentes nodos. Gracias a esto, tareas que antes llevaban horas o días podían realizarse en minutos. Esta forma de trabajar se volvió ideal para empresas que gestionaban grandes volúmenes de información diariamente y necesitaban respuestas rápidas.

Hadoop se compone de varios elementos clave, pero los dos más importantes son **HDFS** y **MapReduce**. HDFS (Hadoop Distributed File System) es un sistema de archivos distribuido que **guarda los datos en diferentes máquinas** y se asegura de que estén disponibles incluso si alguna de ellas falla. Por su parte, MapReduce es un modelo de programación que **divide el trabajo en dos fases**: la fase “Map”, donde los datos se procesan por separado, y la fase “Reduce”, donde se agrupan y se obtiene el resultado final. Todo esto sucede de forma automática, sin que el usuario tenga que preocuparse por los detalles técnicos de la distribución.

Una de las grandes **ventajas de Hadoop** frente a los sistemas tradicionales es su **escalabilidad horizontal**. No hace falta invertir en un gran servidor para

aumentar la capacidad: simplemente se añaden más nodos al sistema. Además, tiene una **alta tolerancia a fallos**, lo que significa que si uno de los equipos deja de funcionar, los datos siguen estando accesibles y el procesamiento no se detiene. Esto es fundamental en entornos donde se trabaja con información crítica o con muchos usuarios simultáneos.

Los usos de Hadoop son muy variados. Por ejemplo, es común encontrarlo en el análisis de **logs de navegación web**, donde se registran millones de visitas al día y se necesita saber qué ha hecho cada usuario, qué páginas ha visitado o cuánto tiempo ha permanecido en ellas. También se utiliza en áreas científicas, como el procesamiento de **datos meteorológicos** para predecir fenómenos naturales o en el estudio de **datos genómicos**, donde se manejan terabytes de información procedente de muestras biológicas.

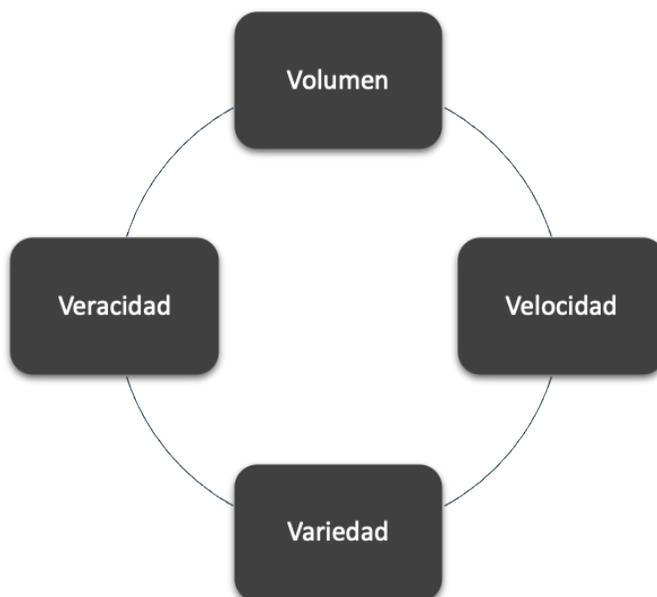
Relación con Big Data	¿Para qué se utiliza?	Ejemplo específico
Hadoop permite almacenar y procesar enormes volúmenes de datos distribuidos en múltiples servidores, algo esencial en entornos Big Data.	Para análisis de datos a gran escala sin necesidad de superordenadores costosos.	CaixaBank lo usa para procesar millones de transacciones diarias y detectar patrones de fraude en tiempo casi real.
Gracias a su arquitectura distribuida, Hadoop permite escalar horizontalmente añadiendo más máquinas según se necesite.	Para almacenar logs, historiales de clientes o sensores IoT que generan datos constantemente.	Renfe podría usar Hadoop para registrar y analizar datos de sensores de trenes (temperatura, frenos, consumo...) en tiempo real para mantenimiento predictivo.
Compatible con otros componentes del ecosistema Big Data como Hive, Pig o HBase.	Para consultas SQL sobre grandes volúmenes de datos o almacenamiento en tablas NoSQL.	INE (Instituto Nacional de Estadística) puede usar Hive sobre Hadoop para hacer análisis demográficos en todos los municipios de España.
Hadoop distribuye el procesamiento (MapReduce) para analizar datos complejos que no caben en una sola máquina.	Para entrenamiento de modelos de IA en grandes volúmenes de datos históricos.	Correos puede usar MapReduce para analizar patrones de entregas retrasadas a partir de millones de envíos.
Hadoop HDFS permite almacenar datos en bruto que luego se procesan con otras herramientas.	Como repositorio central de datos para alimentar dashboards o informes.	RTVE puede usar HDFS para almacenar todas sus emisiones y analizar audiencias en función de contenidos, días y horarios.

Hadoop marcó una diferencia porque **democratizó el acceso al procesamiento masivo de datos**. Gracias a él, muchas organizaciones pequeñas y medianas pudieron acceder a una tecnología que antes solo estaba al alcance de grandes compañías con presupuestos elevados. Hoy en día, aunque han surgido nuevas herramientas que lo complementan o mejoran en ciertos aspectos, Hadoop sigue siendo la base sobre la que se construyó gran parte del ecosistema Big Data actual.

1.4 CARACTERÍSTICAS DEL BIG DATA (4 V'S Y MÁS)

- Volumetría.
- Velocidad.
- Variedad (estructurados/no estructurados).
- Veracidad (calidad del dato).
- Valor del dato.

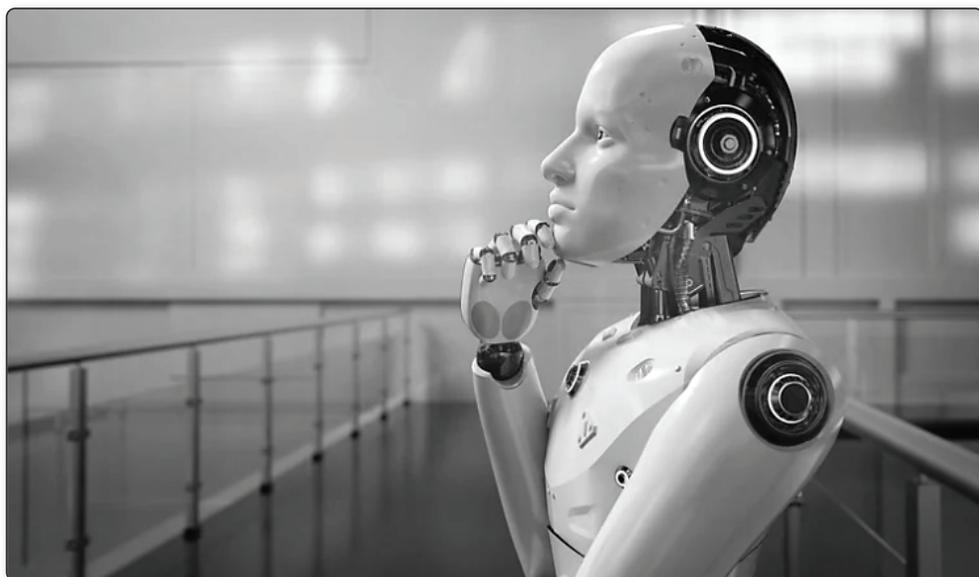
Cuando se habla de Big Data, es muy habitual encontrarse con las llamadas “**4 V's**”, que son una forma sencilla y didáctica de entender qué caracteriza a este tipo de datos y por qué se necesitan herramientas diferentes para gestionarlos. Estas cuatro palabras empiezan por V y resumen las dimensiones clave del fenómeno, aunque con el tiempo se han ido añadiendo algunas más. Vamos a verlas con ejemplos para que se entienda todo de forma clara y cercana.



La primera V es **Volumen**, y tiene que ver con la **cantidad de datos** que se manejan. En el contexto del Big Data, no estamos hablando de gigas o de unos pocos archivos, sino de **terabytes, petabytes o incluso más**. Un buen ejemplo sería el tráfico que gestionan plataformas como **Amazon o YouTube** cada día. Millones de personas ven vídeos, hacen búsquedas, compran productos, dejan opiniones... Toda esa información se acumula y necesita ser almacenada y procesada de forma eficiente. Sin una infraestructura potente, sería prácticamente imposible gestionarlo.

La segunda es **Velocidad**, que se refiere tanto a la **rapidez con la que se generan los datos** como a la rapidez con la que hay que analizarlos. Hoy en día, los datos fluyen continuamente desde distintas fuentes: sensores que miden la temperatura o el tráfico, plataformas de redes sociales que reciben miles de publicaciones por segundo, o sistemas bursátiles que necesitan reaccionar en milisegundos. Cuanto más rápido se procesan esos datos, más útil puede ser la información extraída. Por ejemplo, en la bolsa o en la domótica del hogar, **esperar unos segundos de más puede marcar una gran diferencia**.

La tercera V es **Variedad**, y aquí entra en juego el tipo de datos. Antes se trabajaba casi exclusivamente con datos **estructurados**, que se organizan fácilmente en tablas, como las hojas de cálculo. Pero en Big Data nos encontramos con datos **no estructurados**, como textos, imágenes, audios o vídeos, y también con **semi-estructurados**, como los archivos JSON o XML. Piensa en la cantidad de **opiniones en redes sociales**, correos electrónicos, registros de sensores o grabaciones que una empresa puede tener. Todo eso también es valioso, aunque no se almacene de la misma forma que una tabla de Excel.



La cuarta V es **Veracidad**, que hace referencia a la **calidad y fiabilidad de los datos**. No todo lo que se recoge sirve tal cual: a menudo hay errores, duplicados, inconsistencias o datos incompletos. Por ejemplo, si una tienda online tiene registros duplicados de un mismo pedido, o si un sensor manda una señal errónea porque estaba mal calibrado, eso puede distorsionar el análisis final. Por eso es tan importante filtrar, limpiar y validar los datos antes de analizarlos.

Ejemplo

Datos para ver esto de forma clara, imaginemos un ejemplo concreto: el sistema de emergencias sanitarias de una gran ciudad, como Madrid.

Empecemos por Volumen. Cada día, el 112 recibe cientos de llamadas, muchas de ellas relacionadas con emergencias médicas. A eso se suman los datos que llegan desde ambulancias (GPS, historial del paciente, frecuencia cardíaca en tiempo real), hospitales (registros electrónicos de salud), y centros de coordinación (historiales de incidentes, disponibilidad de recursos). Todo eso se traduce en decenas de miles de registros diarios. En una semana, ya se podrían acumular varios terabytes de información. Este volumen de datos no se puede gestionar con hojas de cálculo o bases de datos tradicionales. Se necesita una infraestructura robusta y escalable, como un data lake o un sistema Hadoop, para almacenar y analizar toda esa información sin que colapse el sistema.

Luego está la Velocidad. En una emergencia, el tiempo lo es todo. El sistema tiene que analizar la información de inmediato: localizar la ambulancia más cercana, saber en qué hospital hay camas disponibles o si se ha producido una acumulación de llamadas en una zona concreta. Todo esto requiere análisis en tiempo real, sin esperar a fin de mes para hacer un informe. Aquí entran en juego tecnologías como Apache Kafka o Spark Streaming, que permiten procesar datos “en caliente” y dar respuestas instantáneas. Si se tarda demasiado, la ayuda podría llegar tarde, y eso puede costar vidas.

La Variedad es también un elemento clave. No se trata solo de números o registros formales. Hay audios de llamadas, mensajes de texto, vídeos de cámaras urbanas, imágenes enviadas desde los móviles de los ciudadanos, historiales médicos en PDF, datos meteorológicos, etc. Toda esta información es útil, pero está en formatos muy distintos. Por eso se necesitan sistemas capaces de entender no solo datos estructurados, como las bases de datos de pacientes, sino también datos no

estructurados o semi-estructurados. Un algoritmo de IA puede, por ejemplo, analizar los mensajes de Twitter geolocalizados para detectar una posible aglomeración o accidente antes incluso de que llegue una llamada al 112.

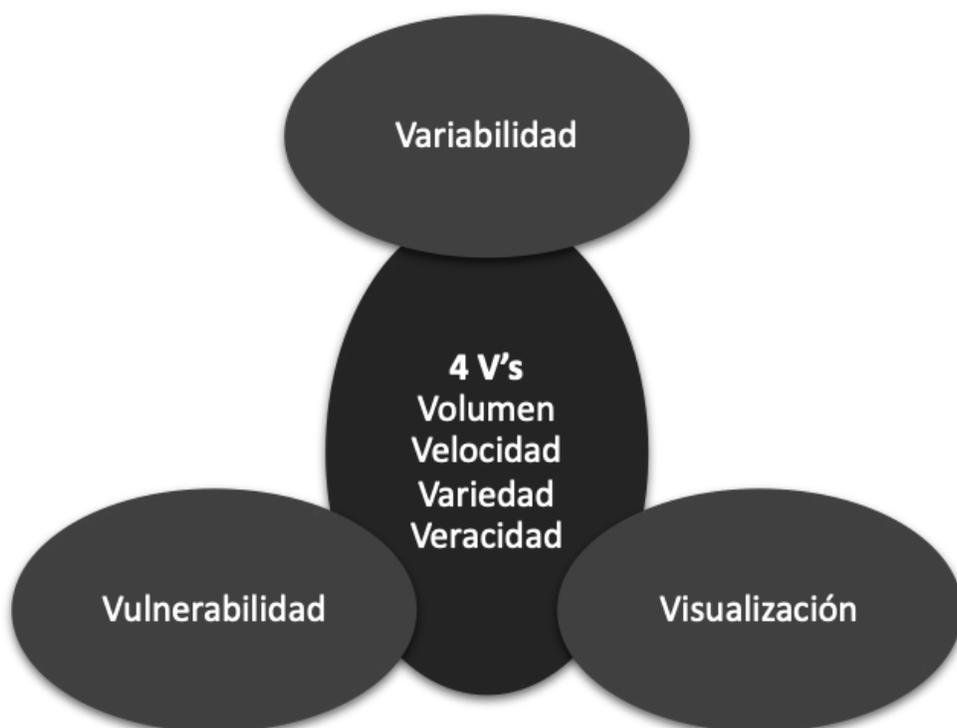
Y finalmente, la Veracidad. En situaciones críticas, no todo lo que se recibe es 100% fiable. Una persona nerviosa puede dar una dirección incorrecta, un sensor puede fallar o los datos pueden estar duplicados. Si el sistema toma decisiones con datos erróneos, se puede enviar una ambulancia al sitio equivocado o malinterpretar la gravedad de un caso. Por eso, antes de procesar los datos, es fundamental aplicar técnicas de limpieza y validación. Se eliminan registros duplicados, se corrigen errores y se comprueban las fuentes. De lo contrario, el sistema puede perder eficiencia o incluso tomar decisiones que generen problemas.

Además de estas cuatro, hay otras V que muchos expertos también consideran clave para entender el ecosistema del Big Data. Una de ellas es la **Variabilidad**, que se refiere a los **cambios frecuentes y rápidos** en los datos. Por ejemplo, una tendencia en redes sociales puede surgir y desaparecer en cuestión de horas, y una campaña de marketing debe adaptarse a esos cambios casi en tiempo real.



Otra V interesante es la **Visualización**. Aunque no se trata de una propiedad de los datos en sí, es fundamental saber **mostrar la información de forma clara y comprensible**, sobre todo para quienes toman decisiones, pero no son técnicos. Un buen gráfico, un dashboard interactivo o una representación visual puede marcar la diferencia entre entender una situación o perderse entre números.

Y por último, no podemos olvidar la **Vulnerabilidad**, ya que el Big Data también conlleva **riesgos de seguridad y privacidad**. Cuanta más información se recoge y más delicada es, más necesario es protegerla. Esto implica pensar en aspectos como el consentimiento del usuario, el almacenamiento seguro o la anonimización de datos personales.



En conjunto, todas estas V nos muestran que el Big Data **no es solo una cuestión de cantidad**, sino también de **velocidad, diversidad, fiabilidad, utilidad, adaptación y responsabilidad**.

Ejemplo

En el caso de la variabilidad, pensemos en una marca de ropa que lanza una nueva colección y empieza a promocionarla en TikTok. Al principio, las interacciones son normales, pero de repente una influencer publica un vídeo viral usando una prenda, y en cuestión de horas el tráfico a la web se dispara, cambian los patrones de búsqueda y los productos más visitados no son los mismos que el día anterior. Sin herramientas de Big Data que capturen y analicen ese cambio casi en tiempo real, la empresa no podría adaptar su stock, sus anuncios ni sus recomendaciones de producto, perdiendo una oportunidad clave de ventas.

Con respecto a la visualización, imaginemos un hospital público que recibe cada día cientos de registros de pacientes, resultados de pruebas, disponibilidad de camas, tiempos de espera, etc. Para que los responsables puedan tomar decisiones rápidas —por ejemplo, reasignar personal a urgencias o abrir una nueva consulta temporal— no basta con tener todos esos datos. Lo que marca la diferencia es disponer de un panel visual claro que muestre en rojo las áreas con más saturación, en verde las que están funcionando bien, y que permita ver tendencias de un vistazo. Ahí es donde herramientas de visualización como dashboards interactivos se vuelven tan esenciales como el análisis mismo.

En cuanto a la vulnerabilidad, un ejemplo muy concreto sería el de una empresa que recopila datos de salud de sus clientes a través de una app de bienestar. Estos datos incluyen frecuencia cardíaca, niveles de estrés, localización y hábitos de sueño. Si esa información no está bien protegida y se produce una filtración, no solo se pone en riesgo la privacidad de miles de personas, sino también la reputación de la empresa. Aquí entra en juego la necesidad de anonimizar los datos, cifrarlos y cumplir con normativas como el RGPD. El Big Data no puede crecer sin una base sólida de confianza y seguridad.

1.5 NUEVOS PARADIGMAS DEL BIG DATA: PROCESOS EN REAL TIME Y CLOUD COMPUTING

En el mundo del Big Data, cada vez se da más importancia a **hacer cosas con los datos justo en el momento en que se generan**, sin tener que esperar horas o días para analizarlos. A esto se le llama **procesamiento en tiempo real**, o “real time”. En lugar de almacenar primero la información y analizarla después, el sistema la interpreta al instante y toma decisiones automáticas. Este enfoque es muy útil en situaciones donde no se puede perder tiempo, como en **alertas bancarias por fraude** (por ejemplo, si se detecta un movimiento extraño en una tarjeta de crédito), o en **recomendaciones en apps móviles**, que cambian según lo que haces en ese momento.

Para que esto sea posible, se utilizan tecnologías que están diseñadas para recibir, interpretar y responder a flujos de datos de manera continua. Algunas de las más conocidas son **Apache Kafka**, que gestiona grandes volúmenes de datos en movimiento; **Spark Streaming**, que permite analizarlos casi en tiempo real; o **Apache Flink**, que es especialmente útil para tareas más complejas y reactivas. Estas herramientas se han convertido en piezas fundamentales para muchas empresas que quieren ofrecer experiencias dinámicas y personalizadas en sus servicios digitales.



Pero además del tiempo real, otro cambio importante que ha revolucionado el Big Data es el uso del **Cloud Computing**, o computación en la nube. Antes, los datos se guardaban y se procesaban en los propios servidores de las empresas, lo que requería mantener infraestructuras costosas, con todo lo que eso implica: espacio físico, mantenimiento, actualizaciones... Hoy en día, gracias a servicios como **Amazon Web Services (AWS)**, **Google Cloud Platform** o **Microsoft Azure**, es posible trabajar directamente en la nube, sin necesidad de tener esos servidores en propiedad.

La ventaja de esta forma de trabajar es que **todo es más flexible y escalable**. Si se necesita más capacidad de almacenamiento o procesamiento, se puede ampliar con solo unos clics. Además, el acceso remoto permite que equipos repartidos por distintas ciudades o países trabajen sobre la misma información sin problemas.

También supone un ahorro importante en costes iniciales, ya que no es necesario invertir tanto en infraestructura física desde el principio.

Cuando se combinan **Big Data y Cloud**, los beneficios se multiplican. Se pueden desarrollar proyectos más ágiles, donde los equipos prueban nuevas ideas sin preocuparse por si los sistemas aguantan o no. Se reduce la dependencia de sistemas locales y se mejora la velocidad para lanzar nuevas soluciones. Por eso muchas aplicaciones y plataformas digitales funcionan directamente sobre la nube.



Un ejemplo muy claro es **Netflix**, que utiliza la nube no solo para almacenar sus contenidos, sino también para analizar los hábitos de millones de usuarios en tiempo real. Gracias a eso, puede recomendar películas, ajustar la calidad del vídeo o incluso decidir qué nuevas series producir. Otro ejemplo son las **ciudades inteligentes (smart cities)**, donde el tráfico urbano se analiza en tiempo real gracias a sensores y cámaras, y los datos se procesan en la nube para gestionar mejor los semáforos, avisar de atascos o mejorar el transporte público.

Este enfoque, que combina velocidad y flexibilidad, **está transformando muchos sectores**, no solo el tecnológico. En salud, por ejemplo, se están empezando a usar dispositivos conectados que monitorizan al paciente en tiempo real y envían los datos directamente a la nube. Esto permite a los profesionales sanitarios reaccionar rápidamente si detectan algún parámetro fuera de lo normal. Lo mismo ocurre en sectores como la logística o la industria: los sensores instalados en vehículos o máquinas pueden avisar de un fallo inminente antes de que ocurra, evitando pérdidas de tiempo y dinero.



También es muy interesante cómo estos nuevos paradigmas han **reducido las barreras de entrada**. Antes, solo las grandes empresas con muchos recursos podían permitirse procesar y analizar grandes volúmenes de datos. Hoy en día, gracias a los servicios en la nube y al procesamiento en tiempo real, **incluso pequeñas empresas o startups** pueden acceder a estas capacidades y ofrecer productos y servicios muy sofisticados. Pueden probar modelos de negocio basados en datos sin tener que montar desde cero un centro de procesamiento o contratar a un ejército de técnicos.

Además, trabajar en la nube y en tiempo real **facilita mucho la escalabilidad**. Es decir, si un proyecto empieza con pocos usuarios y de repente se multiplica la demanda, se puede adaptar el sistema rápidamente sin cambiar toda la arquitectura. Esto es clave, por ejemplo, para aplicaciones móviles que se hacen virales de un día para otro, o para campañas puntuales donde se espera un pico de usuarios, como el Black Friday o eventos deportivos.

Otro aspecto interesante es que, gracias a estas tecnologías, se pueden **automatizar muchas decisiones**. Los sistemas analizan los datos al momento y ejecutan acciones sin intervención humana. Por ejemplo, una plataforma de comercio electrónico puede ajustar automáticamente los precios si detecta que un producto se está vendiendo más rápido de lo habitual. O una empresa de reparto puede reorganizar las rutas de sus transportistas si hay un atasco en una zona determinada.



Todo esto tiene también implicaciones en la **experiencia del usuario**, que se vuelve más fluida, personalizada y adaptada al momento. Si un usuario entra en una app de comida a domicilio y ve recomendaciones basadas en lo que pidió ayer, en la hora del día o incluso en el tiempo que hace, seguramente la sensación será mucho más positiva. Eso es posible porque los sistemas están recogiendo datos y tomando decisiones en tiempo real, apoyándose en servicios en la nube que procesan esa información sin retrasos.

Ejemplo

Plataforma digital de aprendizaje adaptativo Elandra Learning

Elandra Learning es una plataforma digital de aprendizaje adaptativo diseñada para centros educativos de secundaria y formación profesional en zonas rurales de España. Su objetivo es personalizar el proceso educativo de cada estudiante

utilizando tecnologías Big Data, procesamiento en tiempo real y computación en la nube.

¿Cómo funciona? Cada alumno trabaja desde casa o el aula con una tablet conectada a internet. Mientras interactúa con contenidos educativos —como vídeos, ejercicios, simuladores o foros— el sistema va recogiendo continuamente datos:

- Cuánto tarda en responder.
- En qué partes se atasca.
- Cómo navega por los contenidos.
- Qué temas le interesan más...

Todos esos datos se transmiten y procesan al instante gracias a una arquitectura basada en Apache Flink y servicios en la nube de Google Cloud Platform.

Lo más interesante es que Elandra Learning utiliza estos datos para adaptar el contenido de forma inmediata. Si un alumno repite errores en álgebra, por ejemplo, el sistema le propone automáticamente actividades de refuerzo o vídeos explicativos, sin que tenga que pedir ayuda ni esperar a que el profesor lo detecte. Y si progresa más rápido en otra materia, se le desbloquean retos adicionales para mantener su motivación alta.

Además, los profesores cuentan con un dashboard en tiempo real que les muestra alertas sobre los alumnos que están teniendo dificultades, indicadores de avance del grupo, comparativas con el rendimiento esperado, y recomendaciones sobre cómo actuar. Todo eso se actualiza al momento, sin necesidad de revisar manualmente decenas de informes o notas.

Uno de los momentos clave donde Elandra Learning demuestra su potencia es durante las evaluaciones. En vez de un examen único, la plataforma genera cuestionarios personalizados según el nivel de cada alumno y detecta patrones de respuesta para prevenir el plagio o la desconcentración. Incluso adapta el nivel de dificultad según cómo esté respondiendo en tiempo real.

Gracias al uso de la nube, este sistema es accesible desde cualquier dispositivo y no requiere infraestructura especial en los centros educativos. Esto ha permitido llevar educación digital personalizada a lugares donde antes era impensable, reduciendo brechas educativas y ofreciendo una experiencia más justa y adaptativa para todos.

Plataforma de movilidad urbana inteligente ElarinFlow

Imaginemos que una ciudad española como Valencia decide mejorar su sistema de movilidad con una plataforma inteligente llamada ElarinFlow. Esta plataforma combina sensores, cámaras, GPS, análisis de datos y servicios en la nube para tomar decisiones automáticas en tiempo real que afectan directamente a miles de personas cada día.

¿Qué problema quieren resolver?

La ciudad sufre de atascos frecuentes en ciertas zonas y los ciudadanos se quejan de que los semáforos no se adaptan bien al flujo real de tráfico. Además, los autobuses llegan tarde en hora punta y hay poca coordinación entre los diferentes modos de transporte.

¿Qué datos se recogen?

ElarinFlow recopila datos desde múltiples fuentes:

- Cámaras de tráfico que detectan la cantidad de vehículos en tiempo real.
- Sensores en los semáforos que miden cuánto tiempo permanece detenido un coche.
- GPS de los autobuses municipales, que informa de su ubicación y retrasos.
- Información en redes sociales (por ejemplo, quejas de usuarios por atascos).
- Datos meteorológicos en tiempo real (lluvia, calor extremo, etc.).

¿Qué papel juega el procesamiento en tiempo real?

Con tecnologías como Apache Kafka y Flink, los datos fluyen continuamente hacia un sistema de análisis que interpreta al momento lo que ocurre. Si una cámara detecta un atasco que supera cierto umbral, el sistema ajusta automáticamente los semáforos de la zona para priorizar la descongestión. Al mismo tiempo, envía una alerta a los conductores en una app recomendando rutas alternativas, y ajusta el recorrido de los autobuses si hay desvíos necesarios.

¿Y el cloud computing?

Todo este procesamiento ocurre en la nube, concretamente sobre Google Cloud Platform, que permite:

- Escalar automáticamente los servidores cuando hay picos de tráfico (por ejemplo, durante un partido de fútbol).
- Guardar todos los datos históricos para luego analizarlos con más profundidad.
- Compartir en tiempo real la información con distintas áreas del ayuntamiento (tráfico, emergencias, transporte público).
- Mantener un dashboard accesible desde cualquier lugar para visualizar la situación de la ciudad en directo.

¿Qué beneficios consigue la ciudad?

1. Menos atascos:

El sistema reacciona en tiempo real y redistribuye el flujo de vehículos.

2. Más puntualidad en el transporte público:

Los autobuses reciben rutas optimizadas cada pocos segundos.

3. Mejor experiencia del usuario:

La app da recomendaciones personalizadas basadas en el trayecto, el clima o la hora.

4. Ahorro económico:

Se reducen los tiempos muertos, el gasto de combustible y los costes de gestión manual.

5. Sostenibilidad:

Al mejorar el tráfico y evitar recorridos innecesarios, se reducen las emisiones.

Cooperativa tecnológica Sylvaris Agrotech

Sylvaris Agrotech es una cooperativa tecnológica ubicada en La Rioja que trabaja con pequeños y medianos agricultores para optimizar el rendimiento de sus cultivos mediante el uso combinado de Big Data, procesamiento en tiempo real y computación en la nube.

Cada parcela de los agricultores asociados está equipada con sensores de humedad del suelo, estaciones meteorológicas portátiles y cámaras de seguimiento de crecimiento vegetal. Estos dispositivos envían datos constantemente sobre la salud de los cultivos, las condiciones del terreno y el clima local. Los datos se recopilan en tiempo real mediante redes IoT y se envían directamente a la nube, donde se procesan utilizando herramientas como Apache Spark Streaming y servicios de análisis de AWS.

Gracias a esta infraestructura, Sylvaris Agrotech puede, por ejemplo, detectar de inmediato una bajada anómala de humedad en una zona concreta y activar el riego de forma automatizada solo en las áreas afectadas, sin necesidad de regar toda la finca. Del mismo modo, si se detectan condiciones propicias para la aparición de plagas (como una combinación de humedad alta y temperatura cálida), el sistema envía alertas personalizadas al móvil de cada agricultor con medidas preventivas recomendadas.

En la época de vendimia, cuando cada minuto cuenta, Sylvaris Agrotech analiza en tiempo real el estado de maduración de las uvas mediante imágenes multiespectrales recogidas por drones. El análisis se realiza automáticamente en la nube, y en cuestión de minutos se generan mapas de madurez que indican qué parcelas deben recogerse antes. Esto evita pérdidas, mejora la calidad del producto final y reduce costes de forma significativa.

Además, al estar todo alojado en la nube, los agricultores pueden acceder a sus datos y recomendaciones desde cualquier dispositivo, ya estén en el campo, en la bodega o en casa. No tienen que preocuparse por servidores ni por mantenimiento de sistemas, porque Sylvaris Agrotech les ofrece un servicio ágil, adaptable y escalable, incluso si amplían sus hectáreas o se suman nuevos socios a la red.

El procesamiento en tiempo real y el cloud computing han cambiado **la forma de pensar los servicios, los productos y las estrategias empresariales**. Permiten que los datos dejen de ser una fotografía del pasado para convertirse en una herramienta viva, activa y útil en el momento exacto en que se necesita. Y eso, sin duda, está marcando una nueva etapa en la historia del Big Data.

1.6 EL VALOR ESTRATÉGICO DEL DATO EN LAS ORGANIZACIONES MODERNAS

Los datos han pasado de ser un recurso técnico a convertirse en un pilar estratégico en cualquier organización moderna. Durante años, las empresas recogían información que apenas se explotaba. Se almacenaba en bases de datos, se consultaba de vez en cuando y servía, en el mejor de los casos, para hacer informes históricos. Hoy en día, ese enfoque ha cambiado por completo. Las organizaciones que sacan ventaja en sus sectores lo hacen porque han aprendido a usar los datos como una herramienta activa de mejora continua, no como un archivo pasivo.



El verdadero valor del dato está en su capacidad para **orientar decisiones, detectar oportunidades y reducir la incertidumbre**. En un contexto global donde la competencia es intensa y los cambios suceden muy rápido, reaccionar bien ya no es suficiente.

Es necesario anticiparse. Y aquí los datos juegan un papel clave. Por ejemplo, gracias al análisis de patrones de consumo, una empresa puede prever qué productos tendrán más demanda en una temporada concreta, ajustando así su producción y sus campañas de marketing sin tener que ir a ciegas.

En el sector salud, el valor del dato se percibe con claridad. Los hospitales que aplican técnicas de análisis avanzado pueden, por ejemplo, detectar el riesgo de que un paciente desarrolle una enfermedad antes de que aparezcan los síntomas, basándose en historiales clínicos, variables genéticas o estilo de vida. Esto permite intervenir antes y con tratamientos más ajustados. Durante la pandemia, el análisis de datos fue esencial para identificar focos de contagio, organizar recursos hospitalarios y decidir restricciones con base en datos reales, no en suposiciones.

En educación, los centros que utilizan plataformas digitales están empezando a detectar patrones de aprendizaje.

Qué tipo de contenido mejora la retención, en qué momento los estudiantes desconectan o qué factores influyen más en el abandono escolar. Con esa información se pueden diseñar programas formativos más eficaces y personalizados, lo que se traduce en mejores resultados y menos desigualdades.

En comercio, el dato se ha convertido en el eje de la experiencia de cliente. Las grandes plataformas de e-commerce como Amazon, por ejemplo, saben qué productos recomendarte, a qué precio mostrarte una oferta o incluso cuál es el mejor momento para enviarte una notificación, basándose en datos de navegación, historial de compras, clima o ubicación. Esto permite vender más, sí, pero también mejorar la satisfacción del cliente, fidelizarlo y reducir el coste de adquisición.

Otro campo donde los datos son clave es el transporte y la logística. Las empresas de reparto, como Correos Express o SEUR en España, utilizan sistemas basados en datos para calcular rutas óptimas que ahorran combustible y tiempo, reducen emisiones y mejoran los tiempos de entrega. En las ciudades, los ayuntamientos analizan patrones de movilidad para rediseñar líneas de autobús, peatonalizar zonas o ajustar los semáforos en función del tráfico real, no de planes fijos diseñados hace años.



Una tendencia reciente es que **los datos también permiten medir el impacto social y ambiental de las organizaciones**, algo cada vez más demandado por la ciudadanía, la legislación y los propios inversores. Las empresas que son capaces de demostrar con datos reales que están reduciendo su huella de carbono, mejorando la diversidad en sus equipos o generando empleo local, tienen más posibilidades de acceder a financiación o ganar contratos públicos. Aquí entra en juego también la ética del dato: no se trata de recolectar información sin control, sino de usarla con responsabilidad, transparencia y respetando la privacidad.

Ahora bien, tener datos no es lo mismo que usarlos bien. Muchas organizaciones tienen información valiosa pero desorganizada, incompleta o inaccesible para quienes la necesitan. Por eso, el dato no puede depender únicamente del departamento de IT. Su valor estratégico solo se materializa cuando hay una cultura organizativa que lo entiende como un activo compartido, donde todas las áreas –desde finanzas hasta atención al cliente– están implicadas en su recogida, gestión y análisis.

Además, hay un debate interesante que cada vez cobra más fuerza:

¿Qué pasa con las pequeñas empresas o con las organizaciones del sector público que no tienen los mismos recursos que las grandes multinacionales?

Aquí entra en juego la democratización del dato. Plataformas como datos.gob.es o iniciativas como Data For Good permiten que también ayuntamientos, ONGs o cooperativas puedan acceder a información abierta, reutilizarla y sacarle partido para tomar decisiones más informadas. Esto abre una vía muy potente para reducir desigualdades y para que el dato no quede solo en manos de unos pocos.



El dato se ha convertido en un factor estratégico porque permite pasar de la intuición al conocimiento. Pero para que eso funcione, hay que invertir en calidad del dato, en formación para saber interpretarlo y en herramientas que faciliten su análisis y visualización. En este contexto, saber trabajar con datos ya no es solo cosa de analistas o programadores:

Es una competencia transversal que afecta a toda la organización, desde los equipos directivos hasta los operativos. Y es ahí donde reside su verdadero valor.

Aplicación	Sector	Descripción
Predicción de enfermedades crónicas	Seguros	Algoritmos de machine learning analizan historiales médicos y factores de riesgo para anticiparse a enfermedades como la diabetes o el cáncer.
Análisis genético personalizado	Transporte	Big Data permite cruzar millones de muestras genéticas para ofrecer diagnósticos o tratamientos personalizados según el perfil del paciente.
Optimización de recursos hospitalarios	Agricultura	Los hospitales usan análisis predictivo para prever ingresos y liberar recursos críticos como camas UCI o quirófanos.
Detección de fraudes en matrículas	Energía	Sistemas detectan comportamientos sospechosos en la inscripción educativa, como fraude en títulos o duplicación de documentos.
Personalización del aprendizaje	Seguros	Herramientas de análisis ajustan contenidos y ritmos a las necesidades de cada alumno mediante el seguimiento de su progreso digital.
Identificación de abandono escolar	Administración pública	Modelos identifican patrones que alertan del riesgo de que un estudiante abandone, y permiten activar medidas preventivas.
Análisis del comportamiento de compra	Comercio	El análisis de tickets, hábitos de compra y visitas web ayuda a diseñar campañas más efectivas y segmentadas.
Recomendación de productos	Seguros	Sistemas de recomendación muestran productos según historial de navegación y preferencias similares a otros usuarios.
Gestión de inventario en tiempo real	Agricultura	Sensores y datos en tiempo real ajustan los niveles de stock para evitar tanto exceso como faltante en almacenes.
Optimización de rutas logísticas	Energía	Plataformas de logística usan datos de tráfico, tiempo y entregas anteriores para planificar rutas más eficientes.
Predicción del tráfico urbano	Transporte	Cámaras, sensores y datos móviles permiten anticipar atascos y redirigir vehículos en tiempo real.
Monitorización de flotas en tiempo real	Agricultura	Las flotas de autobuses o camiones incorporan sensores para enviar su posición y estado a centros de control logístico.
Gestión de consumo energético	Agricultura	Las empresas eléctricas usan modelos predictivos para gestionar mejor la carga y evitar sobrecostes en horas punta.

Aplicación	Sector	Descripción
Detección de fugas en redes eléctricas	Comercio	Sensores inteligentes detectan fugas o caídas de tensión en infraestructuras eléctricas antes de que se conviertan en averías.
Predicción de picos de demanda	Banca	Algoritmos analizan históricos de consumo y clima para prever picos de demanda y ajustar la producción energética.
Detección de fraudes bancarios	Energía	La banca aplica modelos de IA para identificar transacciones inusuales o patrones típicos de fraude financiero.
Scoring crediticio alternativo	Educación	Nuevos modelos de scoring combinan datos no tradicionales como el comportamiento digital o el historial de pagos informales.
Asistentes virtuales financieros	Agricultura	Los chatbots financieros ofrecen asistencia automática 24/7, resolviendo dudas sobre cuentas, pagos o inversiones.
Detección de fraudes en seguros	Banca	Las aseguradoras usan IA para detectar fraudes al comparar reclamaciones con patrones sospechosos.
Evaluación de riesgo dinámico	Educación	Evaluaciones basadas en IA analizan el historial de cliente, localización y tipo de seguro para ajustar el riesgo.
Procesamiento de imágenes de cultivos	Energía	Drones y satélites recogen imágenes que se analizan para detectar plagas, enfermedades o zonas secas en cultivos.
Control de riego inteligente	Salud	Sistemas de riego automático se activan según los datos de humedad del suelo, clima y tipo de planta.
Predicción de cosechas	Administración pública	Modelos predictivos ayudan a estimar la cantidad de producción en función del clima, el terreno y los cultivos anteriores.
Control de calidad del aire	Banca	Sistemas de sensores miden partículas contaminantes en el aire en tiempo real para informar a autoridades y ciudadanos.
Análisis de impacto ambiental	Medio ambiente	Big Data permite evaluar cómo las acciones humanas o industriales afectan al entorno natural en zonas específicas.
Predicción de incendios forestales	Salud	Modelos predictivos analizan clima, vegetación y actividad humana para anticipar incendios y activar protocolos.
Mejora de servicios públicos	Administración pública	Ayuntamientos usan datos ciudadanos y de sensores urbanos para mejorar la limpieza, iluminación o gestión de residuos.

Aplicación	Sector	Descripción
Análisis de movilidad urbana	Comercio	Las ciudades cruzan datos de tráfico, transporte y movilidad para planificar mejoras urbanas más sostenibles.
Visualización de presupuestos públicos	Seguros	Herramientas de visualización permiten a los ciudadanos explorar presupuestos y partidas públicas con claridad.
Detección de patrones de corrupción	Transporte	Análisis de datos administrativos y financieros detectan redes irregulares o prácticas sospechosas en contratación pública.
Análisis de redes sociales para salud mental	Medio ambiente	Modelos analizan expresiones y actividad en redes sociales para detectar señales de alerta sobre salud mental.
Drones para monitorización agrícola	Comercio	Imágenes aéreas con IA detectan malas prácticas agrícolas o ayudan a monitorizar el crecimiento de cultivos.
Análisis de datos meteorológicos	Energía	Datos del clima se integran con cultivos, riego y logística para mejorar decisiones agrícolas en tiempo real.
Automatización del servicio al cliente	Comercio	Sistemas automáticos responden a preguntas frecuentes en comercios, bancos u organismos públicos, reduciendo tiempos de espera.
Reconocimiento facial en accesos	Seguros	Sistemas de IA validan la identidad del usuario mediante rasgos faciales en accesos a zonas seguras.
Clasificación automática de documentos	Energía	Los documentos se escanean y clasifican automáticamente con IA, reduciendo tareas administrativas manuales.
Análisis de sentimientos ciudadanos	Medio ambiente	Modelos analizan el tono y contenido de mensajes ciudadanos para evaluar preocupaciones sociales o temas emergentes.
Simulación de políticas públicas	Seguros	Gobiernos simulan escenarios con diferentes políticas y sus posibles efectos antes de aplicar medidas reales.
Diagnóstico automático por imagen médica	Educación	IA analiza radiografías o resonancias para detectar enfermedades sin intervención humana directa.
Predicción de accidentes laborales	Transporte	Modelos predicen condiciones de trabajo que aumentan el riesgo de accidentes y permiten tomar medidas preventivas.

Interesante

En una publicación de *AnimalCare* (2025), se detalla el contenido del taller organizado por Veterindustria en colaboración con la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), centrado en las posibilidades de la inteligencia artificial (IA), el Big Data y la digitalización dentro del ámbito de la sanidad animal. La jornada reunió a más de 100 asistentes el 19 de febrero en Madrid.

Durante la apertura, Consuelo Rubio, responsable del Departamento de Medicamentos Veterinarios de la AEMPS, subrayó que estas tecnologías permitirán agilizar procesos regulatorios y de control de medicamentos veterinarios, siempre garantizando la calidad, seguridad y eficacia de los datos utilizados. También remarcó la necesidad de cumplir con las normativas de protección de datos para asegurar una gestión ética de la información.

Desde Veterindustria, Santiago de Andrés, su director general, recalcó el peso estratégico del sector de la sanidad y nutrición animal en España y Europa, dada su aportación a la salud pública, la sostenibilidad ganadera y la seguridad alimentaria. Señaló que España cuenta con 31 plantas de fabricación de medicamentos veterinarios (nueve de ellas dedicadas a productos inmunológicos) y 12 centros especializados en I+D+i, que destinan hasta un 12% de la inversión a investigación.

La sesión de la mañana incluyó intervenciones técnicas por parte de especialistas de la AEMPS. Carmen Pastor introdujo los conceptos básicos de IA y Big Data, José Manuel Simarro explicó cómo se están implantando estas tecnologías en la industria farmacéutica veterinaria, y Ricardo Carapeto junto a Elena Lucas abordaron su papel en la evaluación de riesgos ambientales.

A nivel europeo, Paul Damien Lynn, de la EMA, aportó la visión comunitaria sobre el ciclo de vida de los medicamentos veterinarios y la integración de IA en este proceso.

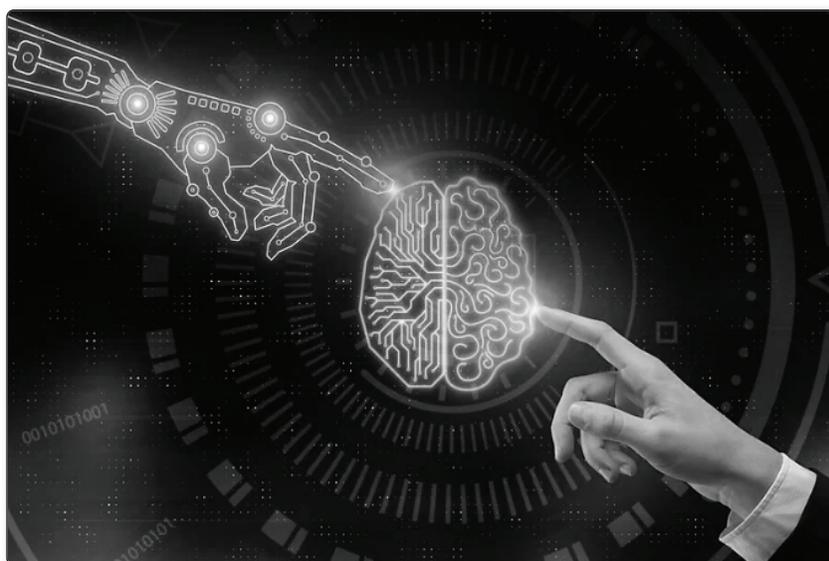
Durante la tarde, Cristina Muñoz analizó ejemplos reales del uso de IA y Big Data en investigación y ensayos clínicos. Finalmente, integrantes del CISA-INIA y el CSIC, como Ana de la Torre, Irene Iglesias y Pablo Ibáñez, detallaron casos concretos de uso en sanidad animal, y Montserrat Portella, de Boehringer Ingelheim, cerró la jornada con una ponencia sobre los retos y oportunidades en la automatización y fabricación de medicamentos veterinarios.

El evento reflejó el creciente interés y compromiso del sector por integrar la transformación digital en todas las fases de la cadena sanitaria veterinaria, desde la investigación hasta la fabricación, con el objetivo de innovar, mejorar procesos y responder de forma más eficiente a los desafíos actuales.

1.7 EVOLUCIÓN Y DISTRIBUCIÓN DEL USO DE INTELIGENCIA ARTIFICIAL Y BIG DATA EN LAS EMPRESAS ESPAÑOLAS

En los últimos años, la adopción de tecnologías emergentes como la inteligencia artificial (IA) y el Big Data ha transformado progresivamente el panorama empresarial en España. Aunque esta transformación no avanza al mismo ritmo en todos los sectores ni en todas las regiones, los datos recogidos por el Instituto Nacional de Estadística (INE) y otros organismos oficiales permiten observar tendencias claras y comprender qué tipo de empresas están liderando esta incorporación tecnológica. Este epígrafe hace un recorrido por el uso de estas herramientas en distintos ámbitos:

Por comunidad autónoma, por actividad económica y por tamaño de empresa, con una atención especial a los casos de las microempresas, que muchas veces quedan fuera del foco de los grandes estudios. Además, se analizan los perfiles profesionales vinculados a estas tecnologías y se compara la situación de España con la de otros países europeos, para ofrecer una visión completa del estado actual y las oportunidades que presenta la transformación digital del tejido productivo.



A lo largo de los siguientes párrafos, se presentarán tablas, gráficos y comentarios que ayudan a contextualizar y reflexionar sobre esta transición tecnológica, mostrando tanto los avances logrados como las brechas que aún persisten. Este análisis permitirá identificar puntos fuertes, debilidades y ámbitos prioritarios de actuación, especialmente en lo que se refiere a la política pública, la capacitación profesional y el fomento de la innovación en las pequeñas y medianas empresas de todo el país.

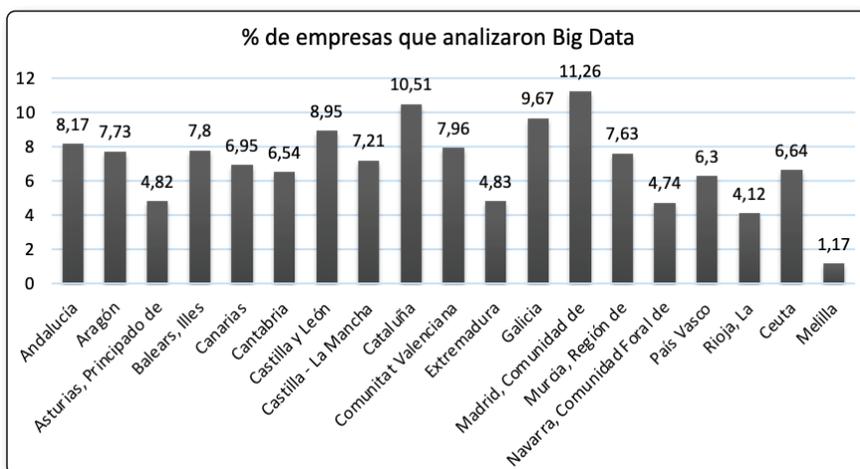
A continuación, se presenta una tabla con el porcentaje de empresas que analizaron Big Data, desglosada por comunidades autónomas en España. Los datos corresponden al primer trimestre del año 2017 y se refieren exclusivamente a empresas con 10 o más empleados, según los registros del Instituto Nacional de Estadística (INE).

Esta información forma parte de un estudio sobre el uso de las Tecnologías de la Información y la Comunicación (TIC) en el tejido empresarial español. El análisis se centra en una de las aplicaciones emergentes más relevantes:

El Big Data. Las cifras reflejan el porcentaje de empresas que han incorporado herramientas de análisis masivo de datos dentro de sus procesos de negocio, según agrupaciones de actividad económica.

Cada porcentaje indica la proporción de empresas que, dentro de cada comunidad autónoma, afirmaron haber utilizado Big Data durante el periodo analizado. Es importante destacar que los códigos al inicio de los literales están vinculados a preguntas concretas del cuestionario original, y que el símbolo “..” señala datos no publicados por razones de secreto estadístico, mientras que “.” indica ausencia de información.

Los datos recogidos son:



Comunidad Autónoma	% de empresas que analizaron Big Data
TOTAL NACIONAL	8,81
Andalucía	8,17
Aragón	7,73
Asturias, Principado de	4,82
Baleares, Illes	7,8
Canarias	6,95
Cantabria	6,54
Castilla y León	8,95
Castilla - La Mancha	7,21
Cataluña	10,51
Comunitat Valenciana	7,96
Extremadura	4,83
Galicia	9,67
Madrid, Comunidad de	11,26
Murcia, Región de	7,63
Navarra, Comunidad Foral de	4,74
País Vasco	6,3
Rioja, La	4,12
Ceuta	6,64
Melilla	1,17

Fuente: Instituto Nacional de Estadística (INE)

Este conjunto de datos permite comprender el nivel de adopción del análisis de datos masivos en cada región y puede servir como base para reflexionar sobre el papel de la innovación tecnológica en el desarrollo económico territorial. También invita a plantearse qué políticas de fomento podrían impulsar esta tendencia en aquellas zonas con menor implantación. ¿Qué barreras encuentran las empresas en regiones como Melilla o La Rioja? ¿Qué factores hacen que Madrid o Cataluña tengan porcentajes tan altos?

Este conjunto de datos ofrece una visión bastante clara de cómo se está adoptando el análisis de Big Data en las distintas comunidades autónomas españolas. Ver los porcentajes permite ir más allá de una simple estadística: ayuda a entender qué regiones están aprovechando el potencial de la analítica avanzada y cuáles todavía tienen camino por recorrer. La diferencia entre comunidades como Madrid o Cataluña, que lideran con cifras superiores al 10%, y otras como Melilla o La Rioja,

que apenas alcanzan el 1% o el 4%, plantea preguntas que merecen explorarse con calma y profundidad.



Estas diferencias pueden reflejar muchas cosas. En primer lugar, el grado de digitalización del tejido empresarial. Madrid y Cataluña cuentan con una mayor concentración de empresas grandes, centros tecnológicos, startups, universidades y hubs de innovación. Todo ese ecosistema favorece la adopción de tecnologías avanzadas, la colaboración entre actores y la generación de talento cualificado. Además, el acceso a financiación, formación y proveedores especializados es mucho más amplio en estas zonas.

En cambio, regiones como Melilla o La Rioja pueden enfrentar barreras importantes:

- **Menor densidad empresarial.**
- **Escasez de personal técnico especializado.**
- **Falta de infraestructuras digitales.**
- **El uso de datos en la toma de decisiones.**

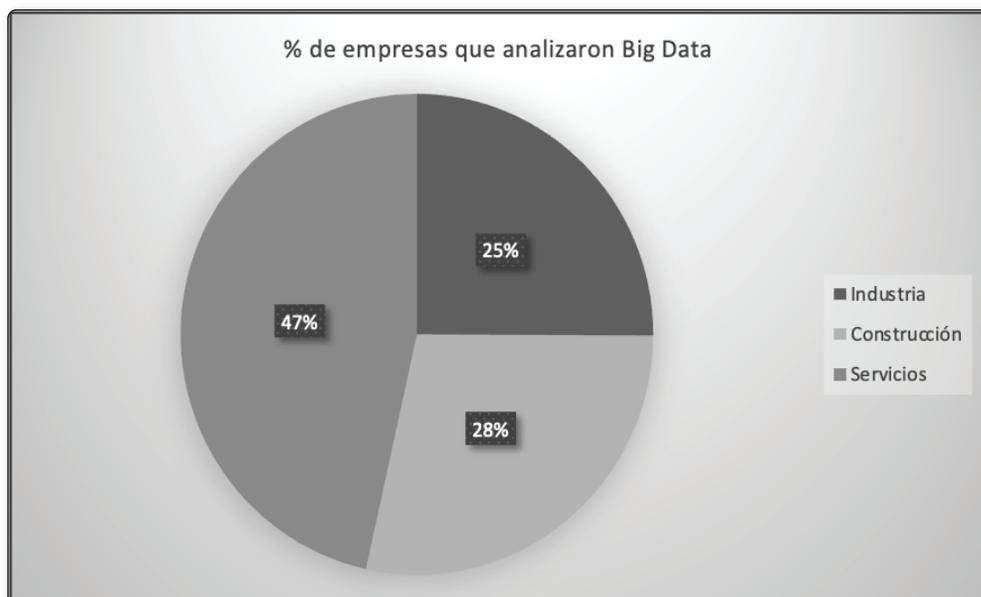
También, hay factores culturales y estructurales que influyen, como el tamaño medio de las empresas, que en muchas zonas rurales tiende a ser más pequeño y con menos recursos para invertir en tecnología.

Este escenario pone sobre la mesa la necesidad de políticas públicas que fomenten la igualdad de oportunidades tecnológicas entre territorios. ¿Qué se puede hacer para que una pyme en Extremadura tenga las mismas opciones que una en Barcelona? Tal vez sea el momento de reforzar los programas de ayudas, crear redes de formación adaptadas a cada región, o incluso establecer centros de innovación compartidos que acerquen estas herramientas a quienes más las necesitan. El desarrollo económico del futuro no depende solo de tener buenos productos o servicios, sino de saber aprovechar los datos para mejorar cada decisión. Y eso debería estar al alcance de todas las regiones.

El siguiente gráfico muestra la distribución del porcentaje de empresas que analizaron Big Data en el primer trimestre de 2017 según el sector económico:

- Industria.
- Construcción.
- Servicios.

A partir de la encuesta sobre el uso de tecnologías de la información y el comercio electrónico, se observa una diferencia bastante clara entre sectores.



Sector	% de empresas que analizaron Big Data
Industria	5,75
Construcción	6,51
Servicios	10,67

Fuente: Instituto Nacional de Estadística (INE)

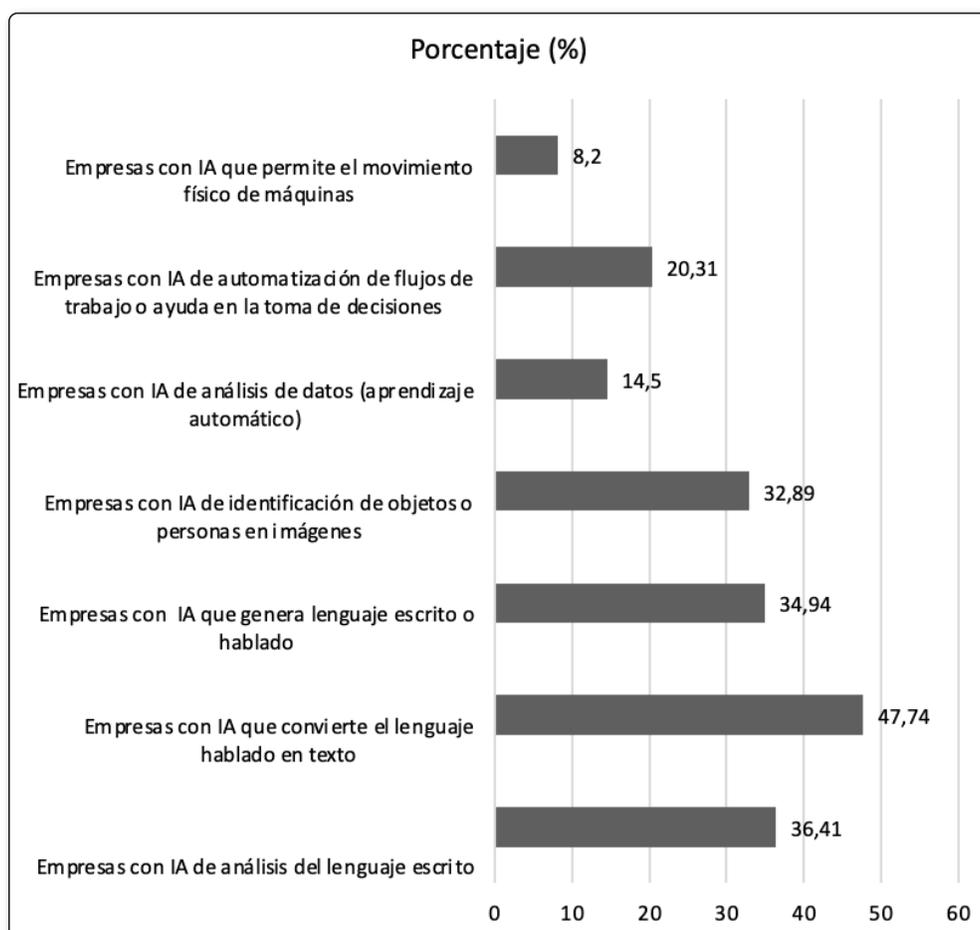
Las empresas del sector servicios lideran con un 10,67 %, lo que indica una mayor adopción del análisis de datos masivos. Esto puede deberse a la necesidad constante de personalizar ofertas, optimizar experiencias del cliente o ajustar campañas de marketing digital. Empresas como bancos, aseguradoras, plataformas de streaming o comercio electrónico suelen operar en este sector y tienen un gran volumen de datos que necesitan procesar en tiempo real.

En segundo lugar, encontramos al sector de la construcción, con un 6,51 %. Aunque a priori pueda parecer un sector más tradicional, cada vez es más habitual el uso de Big Data en la gestión de proyectos, control de costes, análisis de riesgos o mantenimiento predictivo de infraestructuras. Este dato revela que la transformación digital está llegando también a este ámbito, aunque aún con un ritmo más pausado.

Por último, la industria aparece con un 5,75 %. Es un porcentaje más bajo de lo que cabría esperar, sobre todo si se tiene en cuenta el potencial del análisis de datos en fábricas inteligentes (Industria 4.0), control de calidad o gestión de cadena de suministro. Esto podría indicar que muchas pymes industriales todavía no han dado el paso hacia este tipo de soluciones o que existen barreras de acceso tecnológicas o económicas.



El siguiente conjunto de datos recogido por el Instituto Nacional de Estadística muestra el uso de tecnologías de inteligencia artificial (IA) en empresas con menos de 10 empleados en España, durante el primer trimestre de 2022. Este estudio, expresado en porcentajes, permite conocer no solo cuántas de estas pequeñas empresas utilizan algún tipo de tecnología de IA, sino también con qué finalidad concreta lo hacen. Las categorías van desde el análisis del lenguaje escrito, el reconocimiento de voz o de imágenes, hasta usos más avanzados como el aprendizaje automático, la automatización de decisiones o el control de dispositivos físicos. Cada código del informe corresponde a una funcionalidad distinta, lo que facilita un análisis más detallado del tipo de innovación tecnológica que está llegando al tejido empresarial más reducido. Este tipo de información resulta clave para diseñar políticas públicas que fomenten la transformación digital, especialmente en un segmento tan amplio y diverso como el de las microempresas.



Descripción	Porcentaje (%)
Empresas que emplean tecnologías de Inteligencia Artificial (IA)	4,66
Empresas con tecnología IA de análisis del lenguaje escrito	36,41
Empresas con tecnología IA que convierte el lenguaje hablado en texto	47,74
Empresas con tecnología IA que genera lenguaje escrito o hablado	34,94
Empresas con tecnología IA de identificación de objetos o personas en imágenes	32,89
Empresas con tecnología IA de análisis de datos (aprendizaje automático)	14,5
Empresas con tecnología IA de automatización de flujos de trabajo o ayuda en la toma de decisiones	20,31
Empresas con tecnología IA que permite el movimiento físico de máquinas	8,2

Fuente: Instituto Nacional de Estadística (INE)

A continuación, se muestran dos tablas que resumen los usos más comunes de la IA en el ámbito profesional y en el personal, comparando los datos obtenidos en España con los valores globales. Esta información ayuda a comprender mejor el tipo de actividades donde esta tecnología está calando con más fuerza.





Uso de la IA en el ámbito personal	España (%)	Global (%)
Desarrollo de competencias y aprendizaje	48	44
Conocimiento general	46	43
Traducción	36	41
Apoyo para el avance profesional	34	36
Pruebas para conocer qué puede hacer la IA	28	33
Solventar problemas generales/Consejos	28	30
Creación de contenidos creativos	25	27
Recomendación para hobbies	18	19
Productividad y organización	18	16
Amigo virtual y soporte emocional	8	4
Uso de la IA en el ámbito profesional	España (%)	Global (%)
Tareas de escritura	40	41
Estudios, aprendizaje y búsqueda	36	41
Tareas administrativas	34	37
Tareas creativas	31	30
Gestión del conocimiento	26	26
Búsqueda y desarrollo	20	24
Creación de contenidos personalizados	17	14
Programación	17	15
Atención al cliente	11	14
Detección de fraude y cumplimiento normativo	4	7

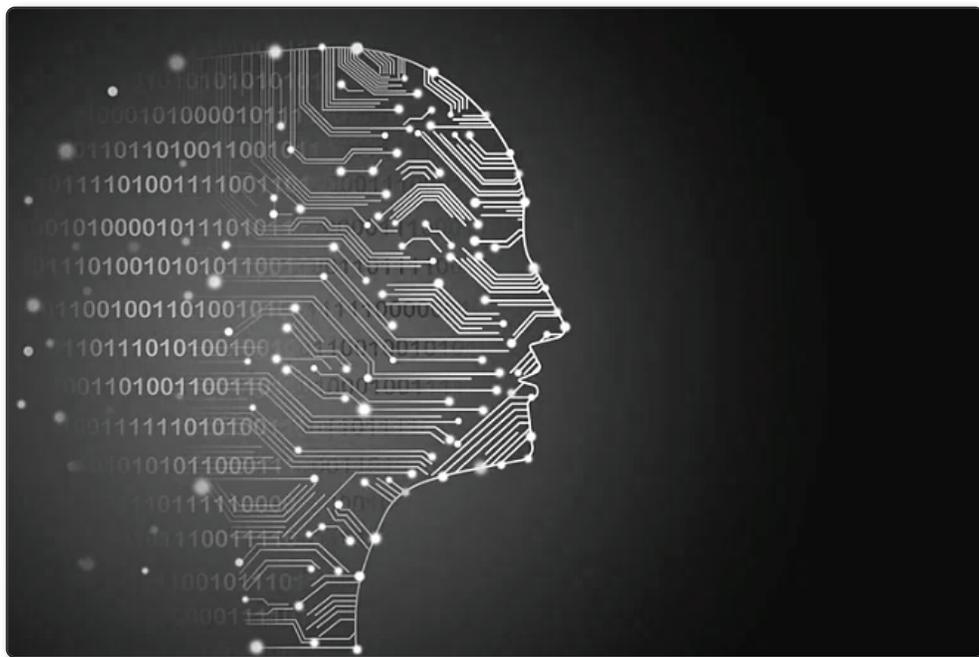
Fuente: Decoding Global Talent 2024

Cuando se trata del uso personal, la inteligencia artificial tiene un papel muy marcado en educación y desarrollo individual. Por ejemplo, el desarrollo de competencias y aprendizaje es el uso más citado (44 %), seguido muy de cerca por el conocimiento general (43 %). Muchas personas recurren a la IA para entender temas, resolver dudas o aprender habilidades nuevas desde casa.

También hay una alta presencia en traducción (41 %), una función muy práctica para quienes viajan, trabajan en entornos internacionales o consumen contenido en otros idiomas. Además, un 36 % la utiliza para recibir apoyo para el avance profesional, ya sea con consejos, preparación de entrevistas o mejora de CVs.

Otro grupo interesante de usos incluye tareas como pruebas para descubrir qué se puede hacer con la IA (33 %), solución de problemas comunes o consejos (30 %), y creación de contenidos creativos como arte, música o escritura, con un 27 %.

En el día a día, algunas personas también la emplean para recomendaciones de hobbies (19 %), organización personal (16 %) o incluso para apoyo emocional o conversación con un “amigo virtual” (4 %), lo cual da que pensar sobre el rol que puede desempeñar la IA en el bienestar.



Según el informe “Uso de inteligencia artificial y big data en las empresas españolas” publicado en 2023 por el Observatorio Nacional de Tecnología y Sociedad (ONTSI), la incorporación de estas tecnologías en el entorno empresarial español continúa en expansión, aunque todavía de forma desigual entre sectores, tamaños de empresa y territorios.



En el año 2022, el uso de inteligencia artificial entre las empresas españolas de más de diez empleados creció hasta alcanzar el 11,8%, lo que representa un incremento de casi cuatro puntos respecto al año anterior. En el caso de las microempresas, el aumento fue más discreto, situándose en torno al 4,6%.

Uno de los usos más frecuentes de la IA es la automatización de procesos y el apoyo en la toma de decisiones, presente en un 46,2% de las empresas que ya usan estas tecnologías. También es común su aplicación para reconocer imágenes o identificar objetos y personas, algo que realizan cerca del 40% de estas compañías.

La adopción de IA no es uniforme. Las empresas del ámbito de la información y la comunicación (41,9%) y las del sector tecnológico (41,3%) lideran con mucha ventaja frente a otros sectores productivos. A nivel territorial, Madrid, Comunidad Valenciana, Aragón y Cataluña son las comunidades que más están incorporando inteligencia artificial, todas por encima de la media nacional.

Aunque ha aumentado el número de empresas que cuentan con especialistas en inteligencia artificial —hasta el 2,3% en 2022—, esta cifra sigue siendo baja en comparación con el total de empresas que utilizan IA. Esto sugiere que muchas organizaciones adoptan herramientas de IA sin contar con perfiles técnicos formados específicamente para ello.

En el contexto europeo, los últimos datos comparables disponibles (2021) situaban a España en una posición intermedia: a la par con la media de la UE-27 (8%) pero todavía lejos de países punteros como Dinamarca (24%), Portugal (17%) o Finlandia (16%).



El análisis de datos masivos también ha experimentado un crecimiento, alcanzando al 13,9% de las empresas españolas con más de diez empleados. En las microempresas, aunque el porcentaje es menor, también crece, situándose en el 3,7%. Son las grandes corporaciones las que más explotan el Big Data, con una implantación del 34,7%. En las medianas la cifra alcanza el 20,8% y en las pequeñas el 11,9%.

Entre las empresas que analizan Big Data, la fuente de información más habitual es la geolocalización de dispositivos móviles, usada por más de la mitad de ellas (55,3%). Le siguen los datos procedentes de redes sociales, empleados por casi el 49%.

De nuevo, los sectores TIC y de información y comunicación son los más avanzados en el uso de Big Data (35,2% y 34,7%, respectivamente). En cuanto a la distribución geográfica, la Comunidad de Madrid encabeza la lista (17%), seguida por Cataluña y La Rioja, ambas con un 16,1%. Por el contrario, Navarra, Castilla y León y Melilla presentan los porcentajes más bajos.

En este campo, España todavía tiene recorrido por delante. En 2021, sólo el 9% de las empresas españolas utilizaban Big Data, cinco puntos por debajo de la media europea. La diferencia es más llamativa si se compara con países como Malta (31%), Países Bajos (27%) o Dinamarca (27%).

¿Cómo están usando la inteligencia artificial y el big data las empresas en España?

El uso de inteligencia artificial (IA) por parte de las empresas españolas sigue creciendo, aunque todavía es algo limitado. En 2022, el 11,8% de las compañías con más de diez trabajadores ya estaba aplicando esta tecnología en sus actividades, lo que supone un avance importante respecto al año anterior. En las microempresas, ese porcentaje es menor, aunque también ha mejorado: un 4,6% frente al 3,5% del año anterior.



Cuando hablamos de big data, los números son algo más altos. Casi el 14% de las empresas medianas o grandes hacen análisis de datos masivos, lo que indica que esta tecnología está algo más extendida. En el caso de las microempresas, el 3,7% la está utilizando, una cifra que también ha subido ligeramente respecto a 2021.

¿Cuáles son las diferencias según el tamaño de la empresa?

El tamaño importa, y mucho, cuando se trata de adoptar nuevas tecnologías. Las grandes compañías son las que más están apostando por la IA: el 41,2% ya la usa, frente al 33,1% del año anterior. Las medianas también han dado un salto importante, pasando del 13,6% al 20%. En las pequeñas, el crecimiento es más discreto, pero constante.

En el terreno del big data, también se nota esta tendencia. Las empresas más grandes han aumentado su uso en más de cinco puntos porcentuales, situándose cerca del 35%. Las medianas han llegado al 20,8% y las pequeñas al 11,9%. Aunque los avances no son espectaculares, muestran una clara evolución hacia una cultura más basada en datos.

¿Qué tipos de tecnologías están usando las empresas?

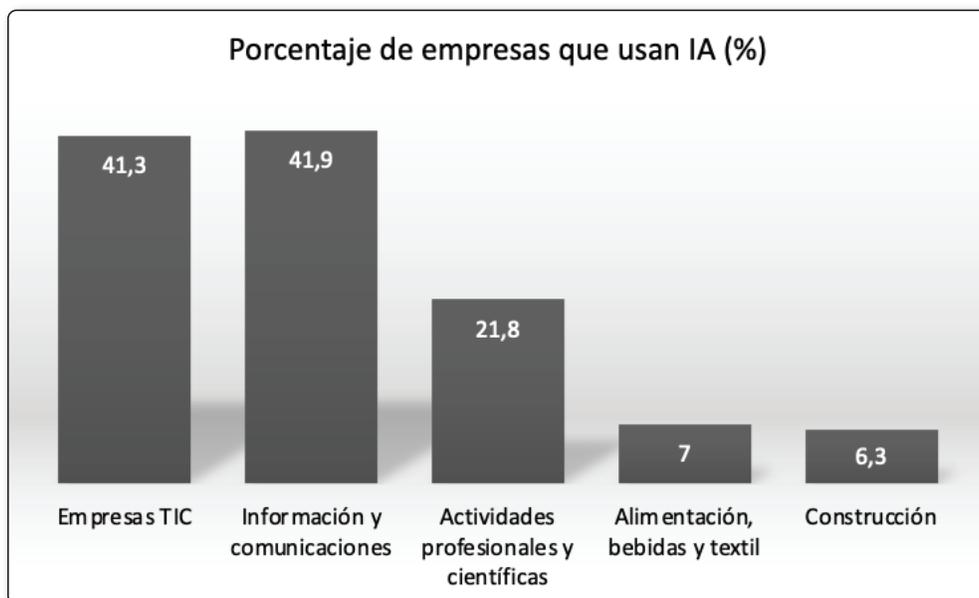
La inteligencia artificial no es una única herramienta, sino un conjunto de tecnologías que pueden tener usos muy diferentes. Por eso, las empresas que ya han empezado a integrarla lo hacen en áreas muy distintas. Las más comunes son:

- Automatización de tareas o apoyo en la toma de decisiones (usada por un 46,2% de las empresas que aplican IA).
- Reconocimiento de imágenes, por ejemplo para identificar objetos o personas (39,7%).
- Aprendizaje automático para analizar datos (33,1%).
- Sistemas que convierten la voz en texto que entiende una máquina (32,4%).

En cuanto al big data, la fuente más habitual para analizar datos masivos es la geolocalización desde móviles u otros dispositivos portátiles. Más de la mitad de las empresas que usan big data trabajan con este tipo de información (55,3%). Los datos provenientes de redes sociales también son muy utilizados (48,6%).

¿Qué sectores lideran el uso de estas tecnologías?

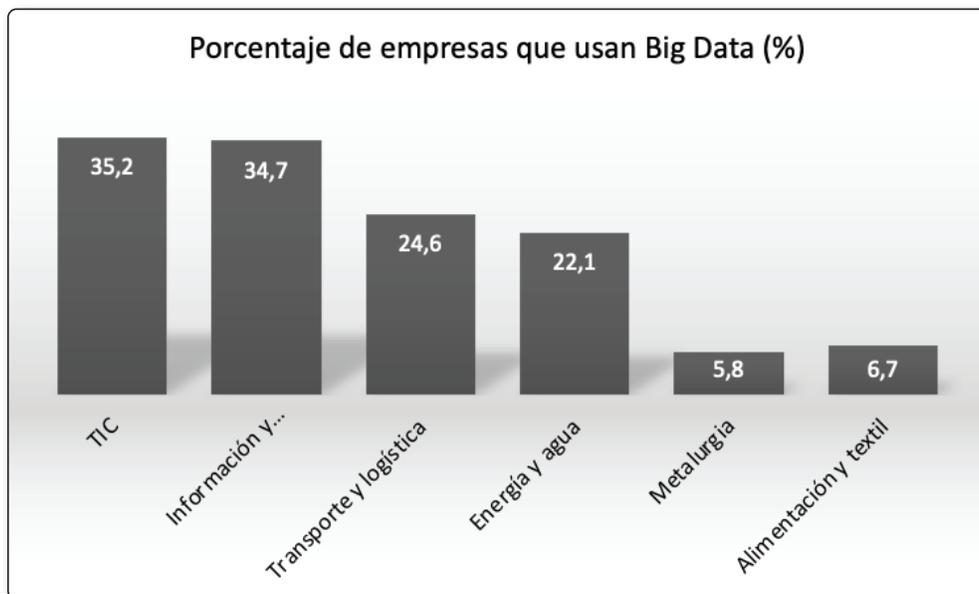
No todos los sectores avanzan al mismo ritmo. La IA, por ejemplo, está mucho más implantada en las empresas tecnológicas o del ámbito de la comunicación. Estas son las cifras más destacadas:



- Empresas TIC: 41,3%.
- Información y comunicaciones: 41,9%.
- Actividades profesionales y científicas: 21,8%.
- Alimentación, bebidas y textil: 7%.
- Construcción: 6,3%.

Con el big data pasa algo parecido.

Los sectores con más implantación son:



- TIC: 35,2%.
- Información y comunicaciones: 34,7%.
- Transporte y logística: 24,6%.
- Energía y agua: 22,1%.
- Metalurgia: 5,8%.
- Alimentación y textil: 6,7%.

Como se puede ver, hay una diferencia de hasta 30 puntos entre los sectores más avanzados y los que apenas han empezado.

¿Y qué pasa en cada comunidad autónoma?

En el mapa de España también hay diferencias notables. Madrid encabeza el uso de inteligencia artificial, con un 16,8% de empresas de más de diez empleados que ya la utilizan. Le siguen la Comunidad Valenciana (14,1%), Aragón (12,7%) y Cataluña (12,3%). En el extremo opuesto están regiones como Castilla y León (7,4%), Ceuta (7%) y Melilla (2,5%).

Si hablamos de big data, el patrón es muy parecido. Madrid vuelve a liderar, con un 17% de implantación. Cataluña y La Rioja le siguen con un 16,1%. Navarra y Castilla y León están bastante por detrás, y Melilla cierra la lista con un escaso 1,6%.

En los últimos años, la expansión de las tecnologías digitales ha traído consigo nuevas profesiones, y entre ellas destacan claramente los perfiles especializados en inteligencia artificial y análisis de datos. Estos avances están transformando la forma en la que se organiza el trabajo, y están obligando tanto a las empresas como a los profesionales a adaptarse, ya sea buscando formación nueva o ajustando sus procesos de contratación para no quedarse atrás en esta nueva realidad tecnológica.



Aunque un número cada vez mayor de empresas españolas (con plantillas de diez o más personas) está usando inteligencia artificial —un 12,6% en 2022—, lo cierto es que muy pocas tienen en plantilla a expertos específicos en IA: apenas un 2,3%. Esto indica que muchas compañías están recurriendo a soluciones ya empaquetadas o a servicios externos, sin necesidad de contar con personal interno especializado. Aun así, esta cifra ha subido casi un punto respecto al año anterior, lo que muestra una tendencia positiva.

Si nos fijamos en el tamaño de las empresas, las diferencias son notables. Mientras que el 10,5% de las grandes empresas ya cuenta con personal experto en IA, este porcentaje baja al 4,4% en las medianas, al 1,6% en las pequeñas y apenas alcanza el 0,1% en las microempresas. El tamaño, como vemos, sigue marcando una gran diferencia en la capacidad para incorporar perfiles especializados.

En cuanto a los sectores económicos que más apuestan por contratar especialistas en inteligencia artificial, destacan claramente el sector TIC, con un 15,5% de empresas que cuentan con este tipo de personal, seguido muy de cerca por el sector de información y comunicaciones (14,5%) y, algo más lejos, el de actividades profesionales, científicas y técnicas (5,5%).

Si hablamos de datos, la proporción de empresas que cuenta con especialistas en este campo es mayor que en el caso de la IA: un 9,8% en 2022. Esta cifra también ha aumentado en comparación con 2021. Al igual que ocurría con la IA, son las empresas grandes las que más incorporan expertos en datos: el 41,7% ya los tiene en plantilla, frente al 20,1% en las medianas y al 7% en las pequeñas.

Por sectores, la tendencia es parecida: las empresas del sector TIC encabezan la lista con un 41,9% que emplea especialistas en datos, seguidas por las del ámbito de información y comunicaciones (40,2%) y las dedicadas a actividades profesionales, científicas y técnicas (12,7%). En cambio, sectores como la construcción o la industria alimentaria están a la cola en este tipo de contrataciones.

Estos datos reflejan una realidad cada vez más clara: el talento especializado en IA y datos está siendo muy demandado, pero aún hay una brecha importante entre la necesidad real y la capacidad de muchas empresas para incorporar este tipo de perfiles, especialmente entre las más pequeñas y en ciertos sectores que aún están lejos de digitalizarse por completo.



En los últimos años, el uso de la inteligencia artificial en las empresas de la Unión Europea ha ido ganando presencia, aunque con importantes diferencias entre países. Cuando se analizan los distintos tipos de tecnologías de IA utilizadas, se observan patrones diversos que reflejan el nivel de madurez digital de cada territorio:

➤ **Análisis del lenguaje escrito**

Uno de los usos más comunes de la IA en las empresas es el análisis del lenguaje escrito. En este campo, países como Portugal, Luxemburgo o Dinamarca destacan con porcentajes que superan el 8%. España, sin embargo, se sitúa en la parte baja de la tabla, con un 2,3% de empresas que aplican esta tecnología, por debajo de la media de la UE27.

➤ **Conversión de lenguaje hablado a formato legible**

Otra aplicación de la IA es transformar lo que se dice en texto que pueda ser procesado por máquinas. Dinamarca lidera esta categoría, seguida de Alemania y Luxemburgo. En este caso, España alcanza el 2,3%, justo en la media europea, lo que indica un desarrollo moderado.

➤ **Generación de lenguaje oral o escrito**

Generar contenido mediante IA, ya sea hablado o escrito, todavía no es muy habitual en Europa. Solo Dinamarca y Eslovenia destacan con cifras superiores al 3%. España, con un 1,4%, empatamos con la media de los países miembros, lo que sugiere que esta tecnología todavía está en fase inicial en el tejido empresarial español.

➤ **Identificación de objetos o personas mediante imágenes**

Las soluciones basadas en visión artificial tienen una acogida desigual. Eslovenia y Finlandia son las más avanzadas, con más del 7% de empresas que usan esta tecnología. España, en este caso, consigue una posición ligeramente por encima de la media con un 3,1%, lo que evidencia cierto avance en este campo concreto.

➤ **Aprendizaje automático**

El machine learning, especialmente aplicado al análisis de datos, presenta una adopción baja en general. Dinamarca vuelve a liderar el ranking con un 8,8%, mientras que España se queda en el 2,3%, por debajo de la media europea (3,7%). Aun así, muestra señales de crecimiento.

➤ **Automatización de procesos y toma de decisiones**

Esta es una de las áreas con mayor potencial práctico. Dinamarca vuelve a encabezar la lista con un impresionante 16,9%, mientras que España se alinea con la media de la UE en un 3%. Esto sugiere que muchas empresas aún están explorando cómo integrar la IA en su operativa diaria.

➤ **Movilidad basada en IA**

El uso de inteligencia artificial para controlar el movimiento físico de dispositivos o máquinas, mediante decisiones autónomas, es muy poco común. En España, solo el 1% de las empresas ha adoptado esta tecnología, que es algo superior a la media europea. Países como Dinamarca y Finlandia están a la cabeza también en esta categoría.

➤ **Uso del big data en Europa**

En lo que respecta al big data, España también se encuentra en una posición intermedia-baja. Solo el 9% de las empresas lo utilizan, un porcentaje inferior al 14,2% de media europea. Países como Malta, Países Bajos y Dinamarca presentan niveles de uso mucho más elevados, por encima del 27%.

➤ **Geolocalización como fuente de big data**

Este tipo de análisis es más habitual en países como Francia, Malta o Bélgica, donde más del 10% de las empresas lo emplean. España, con un 3,2%, se sitúa entre los últimos puestos.

➤ **Big data desde redes sociales**

El análisis de datos generados en redes sociales es otra práctica donde España no destaca: solo un 3% de las empresas lo llevan a cabo, lejos de países líderes como Malta (18%) o los Países Bajos (16%).

➤ **Sensores y dispositivos inteligentes**

Respecto al uso de datos recogidos mediante sensores o dispositivos inteligentes, España también va por detrás. En 2020, apenas un 2,3% de las empresas lo empleaban, frente al 6,8% de media en la UE, y muy lejos de Malta o Finlandia, que superaban el 8%.

Nota

A continuación, se expone un resumen con los **principales datos** extraídos del epígrafe.

ADOPCIÓN DE BIG DATA EN EMPRESAS ESPAÑOLAS



13,9% de empresas de más de 10 empleados usan Big Data (2022).

En microempresas, solo un 3,7%.

Las grandes empresas lo usan en un 34,7% de los casos.

Las fuentes más comunes de datos son:

- **Geolocalización de dispositivos móviles (55,3%).**
- **Redes sociales (48,6%).**

ADOPCIÓN DE IA EN EMPRESAS ESPAÑOLAS



11,8% de empresas de más de 10 empleados usan IA (2022).

En microempresas, el uso alcanza el 4,6%.

Grandes empresas: 41,2% usan IA.

Tecnologías de IA más usadas:

- **Automatización y ayuda en decisiones: 46,2%.**
- **Reconocimiento de imágenes: 39,7%.**
- **Aprendizaje automático: 33,1%.**
- **Conversión de voz a texto: 32,4%.**

SITUACIÓN DE ESPAÑA EN EUROPA



En IA:

- España se sitúa en la media o ligeramente por debajo de la UE en la mayoría de tecnologías.

En Big Data:

- España: 9% de empresas lo usan (2021).
- Media UE: 14,2%.
- Países líderes: Malta, Países Bajos, Dinamarca (27%-31%).

1.8 PRUEBA DE AUTOEVALUACIÓN DEL CAPÍTULO

1. ¿Qué tipo de datos se utilizaban principalmente en el Business Intelligence (BI) tradicional?
 - a) Imágenes y vídeos
 - b) Datos estructurados en tablas
 - c) Datos en tiempo real
 - d) Datos recogidos de redes sociales

2. ¿Cuál de estas herramientas es un sistema distribuido de archivos creado por Google?
 - a) Apache Spark
 - b) Hadoop HDFS
 - c) Google File System (GFS)
 - d) SQL Server

3. ¿Qué empresa desarrolló el modelo MapReduce?

- a) Microsoft
- b) Google
- c) Amazon
- d) IBM

4. ¿Cuál de las siguientes afirmaciones describe mejor la función de MapReduce?

- a) Almacenar datos de forma segura en la nube
- b) Crear dashboards visuales
- c) Procesar datos en paralelo dividiendo tareas entre nodos
- d) Convertir datos no estructurados en estructurados

5. ¿Qué característica del Big Data se relaciona con la rapidez de generación y análisis de los datos?

- a) Variedad
- b) Velocidad
- c) Veracidad
- d) Valor

6. ¿Qué permite Hadoop frente a los sistemas tradicionales?

- a) Usar solo un gran servidor central
- b) Escalar horizontalmente añadiendo más nodos
- c) Evitar el almacenamiento distribuido
- d) Procesar solo datos estructurados

7. ¿Cuál de estos ejemplos refleja claramente un uso de procesamiento en tiempo real?

- a) Análisis mensual de ventas
- b) Alertas de fraude en tarjetas de crédito al momento
- c) Elaboración de informes anuales
- d) Consultas en hojas de cálculo

8. ¿Qué sistema de archivos usa Hadoop?

- a) GFS
- b) HDFS
- c) MapReduce
- d) JSON

9. ¿Qué tipo de datos se considera no estructurado?

- a) Registro de ventas
- b) Texto de opiniones en redes sociales
- c) Tabla de Excel
- d) Archivo CSV

10. ¿Qué ventaja aporta el Cloud Computing al Big Data?

- a) Dificulta el acceso remoto
- b) Permite escalar sin comprar servidores propios
- c) Solo funciona con datos locales
- d) Aumenta los costes iniciales

Respuestas correctas

- 1. b
- 2. c
- 3. b
- 4. c
- 5. b
- 6. b
- 7. b
- 8. b
- 9. b
- 10. b