

# Capacitación de las empresas con la IA: Llega la era de la elección

---



# Índice

La oportunidad de transformar los sectores con la IA.....	1
La IA en los diversos sectores .....	4
Qué deben tener en cuenta los responsables de la toma de decisiones de TI.....	5
Cómo empezar: Desglose de la IA.....	5
Elecciones fundamentales .....	6
Rendimiento .....	6
Seguridad de los datos .....	6
Ampliación de la solución .....	7
Equilibrar los costes y la innovación.....	7
Simplicidad y flexibilidad .....	7
Garantizar la explicabilidad .....	7
Escenarios del mundo real .....	8
Comercio minorista .....	8
Servicios de salud .....	9
Nuestra soluciones .....	10
La IA es para todos: Dell y AMD democratizan la IA .....	10
Colaboración con Hugging Face.....	11
Procesadores AMD EPYC™ .....	11
Aceleradores AMD Instinct™ MI300X .....	11
Plataforma de software de código abierto AMD ROCm™ 6 .....	12
Catálogo de servidores Dell PowerEdge™ .....	12
Resumen.....	13

# La oportunidad de transformar los sectores con la IA

**Hoy en día, la oportunidad para transformar su empresa para el futuro de la innovación está más a su alcance que nunca gracias a la IA. Los datos recopilados de Accenture Vision Technology 2023 muestran que el 98 % de los ejecutivos globales están de acuerdo en que los modelos básicos de IA desempeñarán un papel importante en las estrategias de su organización en los próximos tres a cinco años.<sup>1</sup>**

La IA se ha convertido en una herramienta increíblemente útil para las empresas de campos como el comercio minorista, los servicios de salud y los servicios financieros debido a su capacidad para mejorar la eficiencia en las tareas, impulsar la innovación y mejorar los procesos de toma de decisiones. Sin embargo, a pesar de las ventajas, todavía se percibe una barrera de entrada cuando se trata de integrar la IA debido a algunas ideas equivocadas comunes.



## Necesita un equipo de desarrolladores de IA para empezar:

Aunque los conocimientos en ciencia de datos siguen siendo valiosos para desarrollar soluciones avanzadas de IA y comprender los principios subyacentes, ya no son un requisito previo. Ha habido una proliferación de herramientas de IA fáciles de usar, plataformas como Hugging Face y modelos específicos de tareas que abstraen gran parte de la complejidad involucrada en el desarrollo de soluciones de IA.

## Necesita gastar decenas de millones en hardware para obtener resultados:

Esta idea equivocada socava gravemente la diversidad de los recursos de IA disponibles en la actualidad. Si bien estos recursos comúnmente conocidos suelen ser potentes y disponer de soporte adecuado, es posible que no siempre sean la opción más adecuada o rentable para todas las empresas.

## Hay que trabajar incansablemente para adquirir aceleradores:

Aunque los aceleradores son excelentes para las grandes cargas de trabajo de IA, es posible que las empresas no necesiten tanta potencia de computación para sus aplicaciones de IA. Esperar un periodo de tiempo excesivamente largo para acceder a los aceleradores líderes del mercado tampoco es demasiado realista. En muchos casos, las CPU optimizadas para IA pueden ofrecer el rendimiento y la eficiencia necesarios para generar decisiones y análisis asistidos por IA en tiempo real, y son una solución mucho más rentable y adaptable.

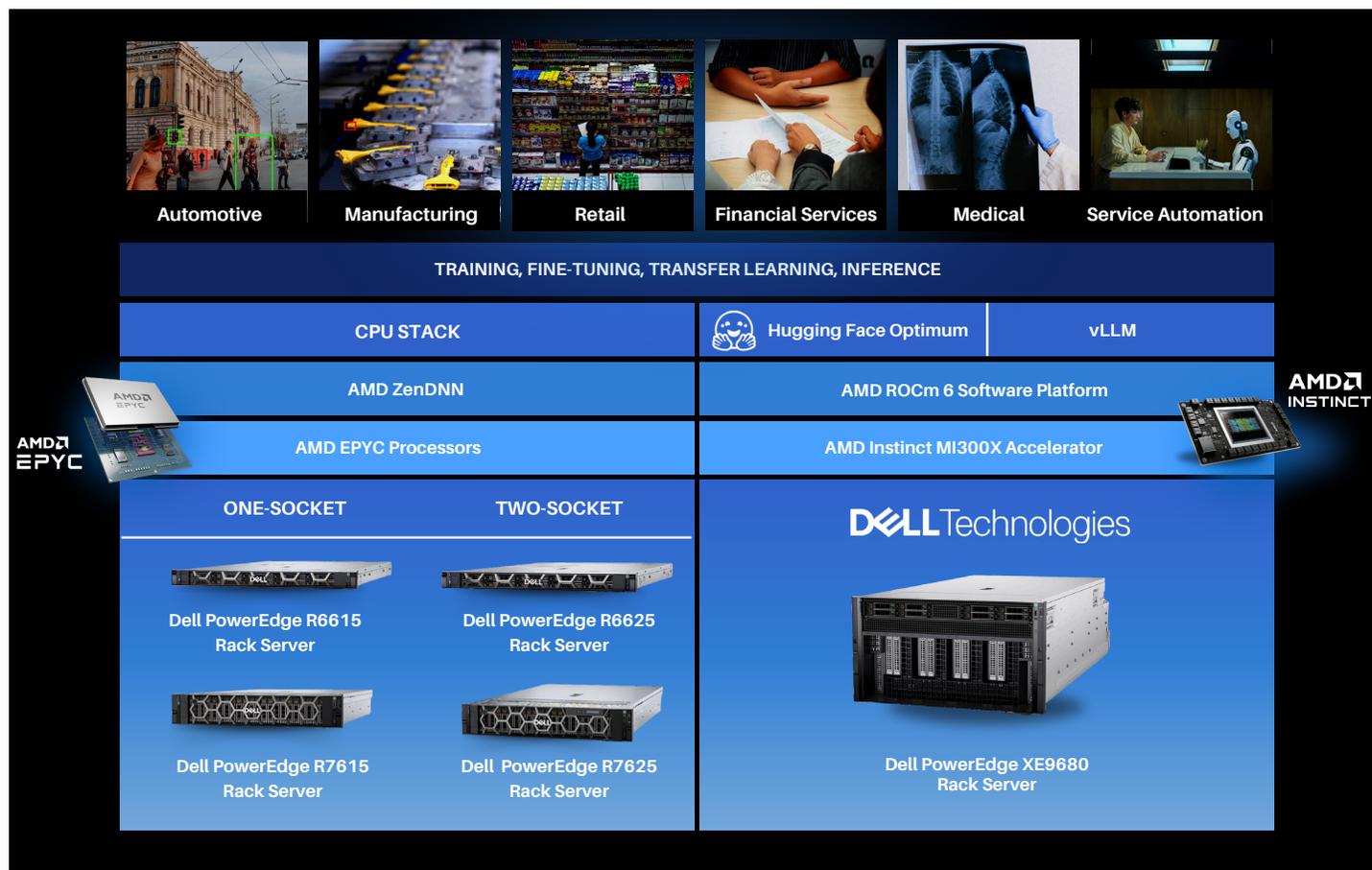
<sup>1</sup> Accenture, 30 de marzo de 2023, "Accenture Technology Vision 2023: Generative AI to Usher in a Bold New Future for Business, Merging Physical and Digital Worlds", <https://newsroom.accenture.com/news/2023/accenture-technology-vision-2023-generative-ai-to-usher-in-a-bold-new-future-for-business-merging-physical-and-digital-worlds>



Afortunadamente, el panorama de la IA está evolucionando. **Dell** y **AMD** se han asociado para romper estos mitos haciendo que las tecnologías y herramientas de IA sean accesibles para una variedad más amplia de usuarios con una infraestructura integral diseñada para satisfacer las demandas de IA de hoy en día.

Puede comenzar con un modelo ya optimizado, una pila de software fiable y un sistema de hardware versátil, todos ellos disponibles abiertamente gracias a la asociación entre Dell y AMD. Para aprovechar la IA, ya no es necesario tener acceso a aceleradores cada vez más escasos, a un grupo sustancial de ingenieros de IA cualificados o a recursos para implementar clústeres de cloud masivos.

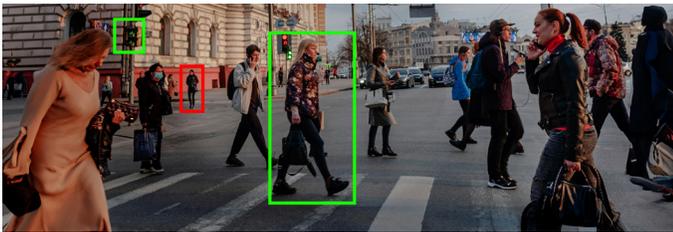
Lo que la colaboración entre **Dell** y **AMD** ofrece es un ecosistema unificado de hardware y software diseñado para permitir a los desarrolladores crear soluciones de IA integrales que incorporen el aprendizaje por transferencia, el ajuste preciso y la inferencia de manera fácil y eficiente. Con el respaldo de **Hugging Face**, ahora contamos con una cartera cada vez mayor de modelos que se ejecutan en servidores Dell PowerEdge con procesadores AMD EPYC™ o aceleradores AMD Instinct™ MI300X para que los desarrolladores puedan realizar ajustes precisos, aplicar el aprendizaje por transferencia e implementar modelos para la inferencia. Las inversiones en AMD ROCm™ y AMD ZenDNN™, así como las asociaciones con los marcos PyTorch, Tensorflow y ONNX Runtime, son los facilitadores fundamentales de los desarrolladores de IA aplicada que experimentan la democratización de la IA. En el siguiente diagrama de bloques se detallan los componentes que componen el ecosistema de IA unificado de Dell y AMD.



# La IA en los diversos sectores

Con la diversificación de los recursos y el énfasis en la innovación de código abierto, la IA está migrando a muchos sectores diferentes, como los del servicio de atención al cliente, las finanzas y la banca, los servicios de salud y el comercio minorista, por nombrar algunos. Sin embargo, en todos estos sectores, la IA permite a las organizaciones aprovechar el potencial de sus propios datos y reinventar sus flujos de trabajo de IA a través de las siguientes capacidades clave: análisis de datos, automatización, personalización y análisis predictivo. Además, las bibliotecas ROCm y ZenDNN de AMD aceleran estos flujos de trabajo de IA para ofrecer resultados casi en tiempo real.

Eche un vistazo a continuación y descubra cómo influye exactamente la IA en varios sectores.



## Automoción

La IA se utiliza para la detección de objetos, la ayuda para la permanencia dentro del carril y la toma de decisiones en vehículos autónomos. La IA también puede predecir cuándo es probable que falle un componente del vehículo, lo que permite un mantenimiento proactivo y reduce el tiempo de inactividad.



## Fabricación e industria

La IA se puede utilizar en la fabricación y la industria para el mantenimiento predictivo, el control de calidad, la optimización de procesos y la administración de la cadena de suministro, lo que mejora la eficiencia y reduce el tiempo de inactividad.



## Comercio minorista

La IA puede analizar el comportamiento de los clientes para ofrecer recomendaciones de productos personalizadas, mejorando así el compromiso del cliente y las ventas. También puede optimizar los niveles de inventario al predecir la demanda y minimizar el exceso de existencias o el desabastecimiento.



## Servicios financieros

La IA se puede utilizar en las finanzas y la banca para la detección de fraudes, la evaluación del riesgo, el servicio de atención al cliente y el análisis de inversiones, lo que mejora la seguridad y permite una toma de decisiones mejor fundamentada.



## Medicina

La IA se puede utilizar en el sector de los servicios de salud para una variedad de aplicaciones, como el análisis de imágenes médicas, el diagnóstico de enfermedades, la planificación personalizada del tratamiento y el descubrimiento de fármacos, lo que permite mejorar los resultados de los pacientes y reducir los costes.



## Automatización de servicios

Los chatbots basados en IA pueden gestionar las consultas de los clientes y proporcionar asistencia, lo que reduce la necesidad de intervención humana. La IA también puede automatizar tareas repetitivas, como la introducción de datos o el procesamiento de documentos, lo que mejora la eficiencia y reduce los errores.

# Qué deben tener en cuenta los responsables de la toma de decisiones de TI

## CÓMO EMPEZAR: DESGLOSE DE LA IA

Antes de examinar estos casos de uso, analicemos en mayor detalle el ciclo de vida de la IA. El ciclo de vida de la IA (Inteligencia Artificial) hace referencia a las etapas involucradas en el desarrollo, la implementación y el mantenimiento de un sistema de IA. Aunque las metodologías y la terminología específicas pueden variar, un ciclo de vida típico de la IA siempre incluye entrenamiento e inferencia de modelos.

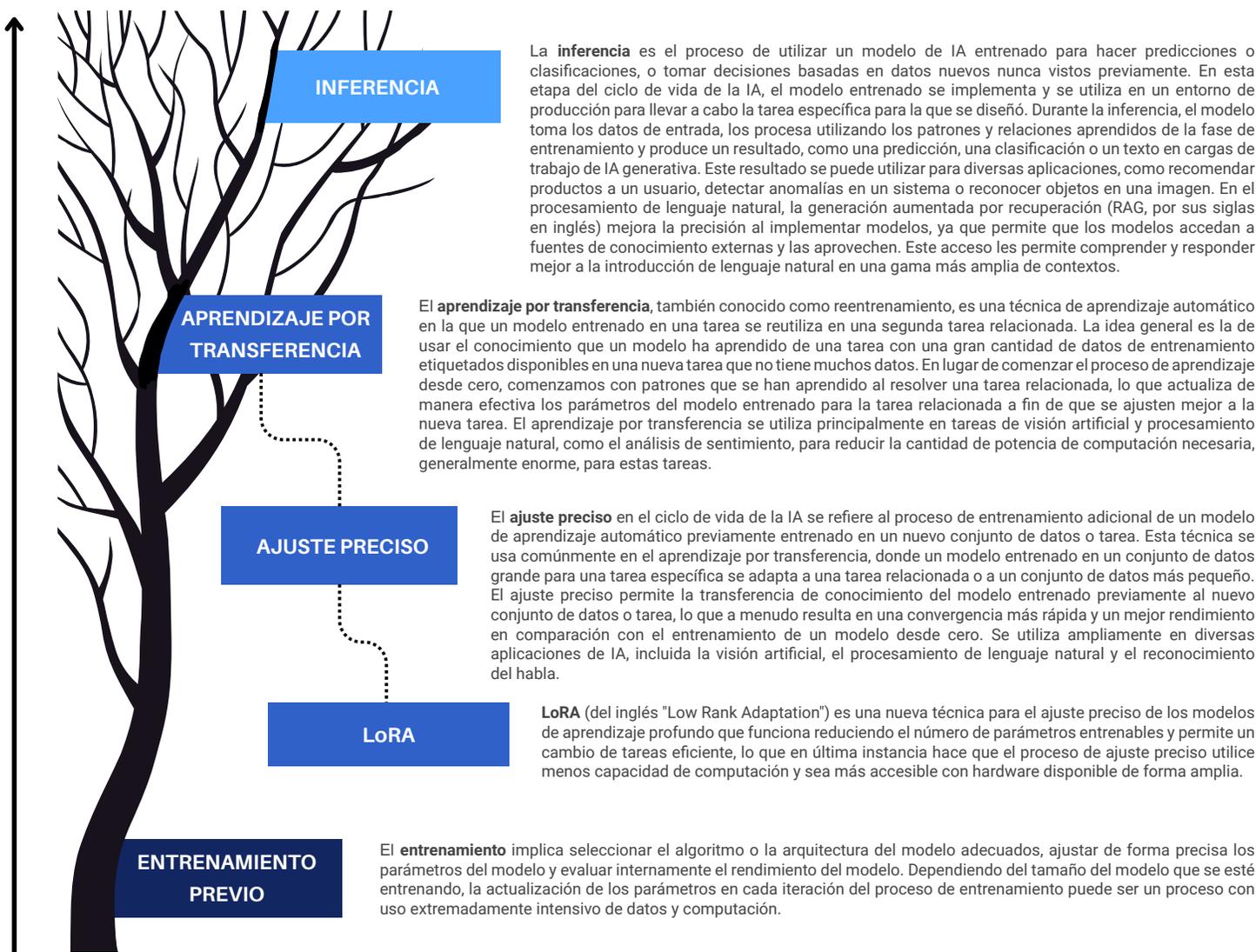


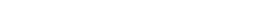
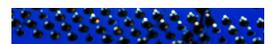
Figura 1: El ciclo de vida de la IA



## ELECCIONES FUNDAMENTALES

### | Rendimiento

En muchas de estas aplicaciones del mundo real, la toma de decisiones en tiempo real o casi en tiempo real es fundamental para el éxito. Por ejemplo, las actividades fraudulentas en las transacciones financieras o las reclamaciones de seguros deben identificarse rápidamente para evitar pérdidas financieras y proteger los activos de las empresas. En un escenario de fabricación, los fallos en la línea de montaje o las condiciones de la fábrica deben supervisarse dinámicamente para garantizar la calidad. El procesador que gestiona la carga de trabajo de inferencia debe optimizarse efectivamente para procesar los flujos de datos entrantes de manera rápida y eficiente. La combinación de los servidores Dell PowerEdge y los procesadores AMD EPYC es muy versátil, ideal para gestionar cargas de trabajo de inferencia en el perímetro, así como tareas relacionadas con la informática de alto rendimiento, el cloud computing y el análisis de Big Data.



### | Seguridad de los datos

La **seguridad de los datos** es crucial para el éxito de los sistemas de IA, especialmente los que aprovechan la IA generativa, y es una preocupación importante para los líderes tecnológicos que pretenden incorporar la IA a sus operaciones. Los sistemas de IA suelen basarse en cantidades masivas de datos, que pueden incluir información delicada y confidencial, como datos personales, datos financieros o información privada. Proteger estos datos es fundamental para evitar el acceso no autorizado o el robo de datos, así como para garantizar la precisión, fiabilidad y coherencia de los modelos y predicciones de IA.

La **computación confidencial** es una tecnología que facilita el procesamiento de datos en un enclave seguro, protegiéndolos del acceso no autorizado o la manipulación por parte de personas no autorizadas, incluidos el proveedor de cloud y otros usuarios.<sup>2</sup> Para aislar los datos durante el procesamiento se utilizan el cifrado y otras medidas de seguridad. AMD Infinity Guard, una colección de sofisticadas características de seguridad integradas en los procesadores AMD EPYC, admite la computación confidencial mediante el empleo de Secure Encrypted Virtualization (SEV), que cifra las máquinas virtuales (VM) mediante una clave que solo el procesador conoce. Estos servicios tienen como objetivo proporcionar entornos de ejecución de confianza basados en hardware mediante AMD SEV-Secure Nested Paging (SEV-SNP), que mejora las protecciones de los sistemas invitados para ayudar en la defensa frente a las amenazas externas.

El **aprendizaje federado** es otro método para mantener la seguridad de los datos. Entrena un modelo central a través de dispositivos o servidores descentralizados.<sup>3</sup> En lugar de transferir todos los datos a una ubicación central, cada dispositivo entrena el modelo localmente y solo se comparten las actualizaciones del modelo. Este enfoque preserva la privacidad y permite el aprendizaje colaborativo sin compartir los datos sin procesar. La plataforma de IA federada de Dell Technologies permite la ejecución de procesos de computación, IA y algoritmos de ML en conjuntos de datos en el perímetro de la red cuando se recopilan, y solo se comparten modelos matemáticos, metadatos y resultados de consultas a través de la red con otros dispositivos del perímetro, centros de datos o el cloud. Este intercambio mejora los resultados al permitir la extracción casi en tiempo real de información práctica de grandes conjuntos de datos distribuidos sin revelar los datos ni la propiedad intelectual.

<sup>2</sup> Advanced Micro Devices, Inc., 2023, 30 de agosto, "AMD shares the technical details of technology Powering Innovative Confidential Computing Leadership Cloud Offerings", <https://www.AMD.com/en/newsroom/press-releases/2023-8-30-AMD-shares-the-technical-details-of-technology-pow.html>  
Advanced Micro Devices, Inc., 2021, informe sobre la solución "Data Center Solutions, Confidential Computing", <https://www.AMD.com/content/dam/AMD/en/documents/EPYC-business-docs/solution-briefs/confidential-computing-solution-brief.pdf>

<sup>3</sup> Analytics Vidhya, diciembre de 2023, "Federated Learning: A Beginner's Guide", <https://www.analyticsvidhya.com/blog/2021/05/federated-learning-a-beginners-guide/#:~:text=Federated%20learning%20works%20by%20training,learning%20without%20sharing%20raw%20data>  
Dell Technologies, 2021, informe sobre la solución "A federated learning platform for real-time artificial intelligence", <https://www.Delltechnologies.com/asset/en-us/solutions/industry-solutions/industry-market/dt-sb-analytics-anywhere.pdf>

## AMPLIACIÓN DE LA SOLUCIÓN

### | Equilibrar los costes y la innovación

Lograr el equilibrio adecuado entre los costes y la innovación garantiza que las soluciones de IA no solo sean viables desde el punto de vista financiero, sino que también tengan un impacto, y que generen un valor real tanto para las empresas como para los usuarios. Un componente clave para encontrar este equilibrio radica en identificar un hardware que resuelva sus casos de uso y, además, se integre fácilmente en la infraestructura existente. En el mercado moderno de hardware de IA, el aumento de la demanda de aceleradores en diversos sectores, además de las limitaciones de capacidad de producción, los desafíos logísticos y la escasez de semiconductores, están contribuyendo a la escasez de aceleradores.

Sin embargo, las CPU ya son un componente estándar en la mayoría de los centros de datos, lo que hace que la integración sea más simple y rentable en comparación con la adición de hardware de acelerador completamente nuevo. Las CPU optimizadas para IA pueden aprovechar el software y las herramientas existentes, lo que reduce la necesidad de un reequipamiento o reentrenamiento exhaustivos. Las CPU también ofrecen una mayor flexibilidad y eficiencia para una amplia gama de tareas que van más allá de la IA, lo que permite un uso más versátil de los recursos dentro del centro de datos. Actualizar el centro de datos con servidores Dell PowerEdge que ejecuten procesadores AMD EPYC permite cumplir con las cargas de trabajo existentes, a la vez que se mantienen preparados para adaptarse a los avances hacia una mayor innovación y eficiencia impulsadas por la IA.

### | Simplicidad y flexibilidad

La simplicidad y la flexibilidad de su sistema de IA son esenciales para crear soluciones de IA que sean eficaces, adaptables y ampliables a largo plazo. Tener acceso a una suite de marcos de software y optimizaciones que complementan su hardware mejora el rendimiento sin dedicar tiempo y esfuerzo adicionales a la integración multiplataforma. Estas cualidades son especialmente importantes para afrontar cargas de trabajo de IA mixtas, que implican una combinación de diferentes tipos de tareas de IA, como el entrenamiento, la inferencia y el procesamiento de datos.

AMD y Dell Technologies afrontan las cargas de trabajo mixtas de IA mediante una combinación de soluciones de hardware y software. Los procesadores AMD EPYC proporcionan potencia de informática de alto rendimiento, con funciones como los multiprocesos simultáneos (SMT, por sus siglas en inglés) y un alto número de núcleos, lo que permite un procesamiento en paralelo eficiente para cargas de trabajo de IA. Estos procesadores están optimizados para tareas de IA y ofrecen un sólido rendimiento tanto para cargas de trabajo de entrenamiento como de inferencia. Los servidores Dell PowerEdge, equipados con procesadores AMD EPYC, proporcionan una plataforma ampliable y flexible para implementar cargas de trabajo de IA. Además, la suite Dell OpenManage Software ofrece herramientas de gestión para optimizar la asignación de recursos y la supervisión del rendimiento de las cargas de trabajo mixtas de IA.

AMD también ofrece Unified Inference Frontend (UIF), que aprovecha las versiones de rendimiento mejorado de cada una de las pilas de software actuales y se basa en la biblioteca AMD ZenDNN para los procesadores AMD EPYC, la pila AMD ROCm de código abierto para los aceleradores AMD Instinct y una pila de software para los sistemas en chip (SoC, por sus siglas en inglés) adaptables de AMD. AMD ROCm también se ha diseñado para funcionar con una amplia variedad de CPU y aceleradores de AMD, incluidos productos tanto para uso profesional como para consumidores.

### | Garantizar la explicabilidad

La **IA explicable** desempeña un papel fundamental a la hora de garantizar la transparencia, la fiabilidad y la eficacia en las aplicaciones de inteligencia artificial. La IA explicable proporciona información sobre la forma en que los modelos de IA toman decisiones, y arroja luz sobre los factores subyacentes y los procesos de razonamiento. Esta transparencia es crucial para ganarse la confianza de las partes interesadas, especialmente en ámbitos delicados, como los servicios de salud, las finanzas y el sistema judicial, donde las decisiones afectan directamente a la vida de las personas.

Los sistemas de IA "**human-in-the-loop**" (con intervención humana) aprovechan la inteligencia humana para mejorar el rendimiento de la IA y mitigar los sesgos algorítmicos. Al integrar la supervisión humana, estos sistemas pueden gestionar situaciones complejas y ambiguas de manera más efectiva, lo que garantiza que las soluciones de IA se alineen con las normas éticas y sociales. Además, la intervención humana permite el perfeccionamiento y la adaptación continuos de los modelos de IA en función de la información del mundo real, lo que fomenta la mejora iterativa y la fiabilidad a largo plazo. Estos enfoques son esenciales para crear sistemas de IA responsables, fiables e inclusivos que sirvan a los mejores intereses de la sociedad.

# Escenarios del mundo real

Scalers AI colaboró con Dell y AMD para mostrar las capacidades de los servidores Dell PowerEdge equipados con procesadores AMD. Descubra cómo se aprovechan estas tecnologías para la formación, el aprendizaje por transferencia y la inferencia en escenarios de comercio minorista y servicios de salud.

## COMERCIO MINORISTA

Scalers AI creó la solución de referencia de gestión de inventario para comercios minoristas, un sistema diseñado para supervisar y gestionar los niveles de existencias en los estantes de los comercios minoristas a través de la implementación de un modelo de IA de detección de objetos. Esta solución de referencia aprovecha el modelo SSD\_MobileNet\_V2 para identificar y reconocer productos en los estantes de las tiendas, lo que, en última instancia, permite recuentos automáticos de inventario y una supervisión precisa de los niveles de existencias. El modelo se sometió al aprendizaje por transferencia utilizando el conjunto de datos de imágenes SKU110K, que contiene 23 000 imágenes de Roboflow. Al aprovechar la visión artificial y los algoritmos de aprendizaje automático, el sistema puede detectar cuándo los artículos se están agotando o están desabastecidos, y proporciona alertas al personal de la tienda para reabastecerlos o reponerlos a tiempo.

Esta solución utiliza el servidor Dell PowerEdge R7615 con el procesador AMD EPYC 9354P de 32 núcleos.

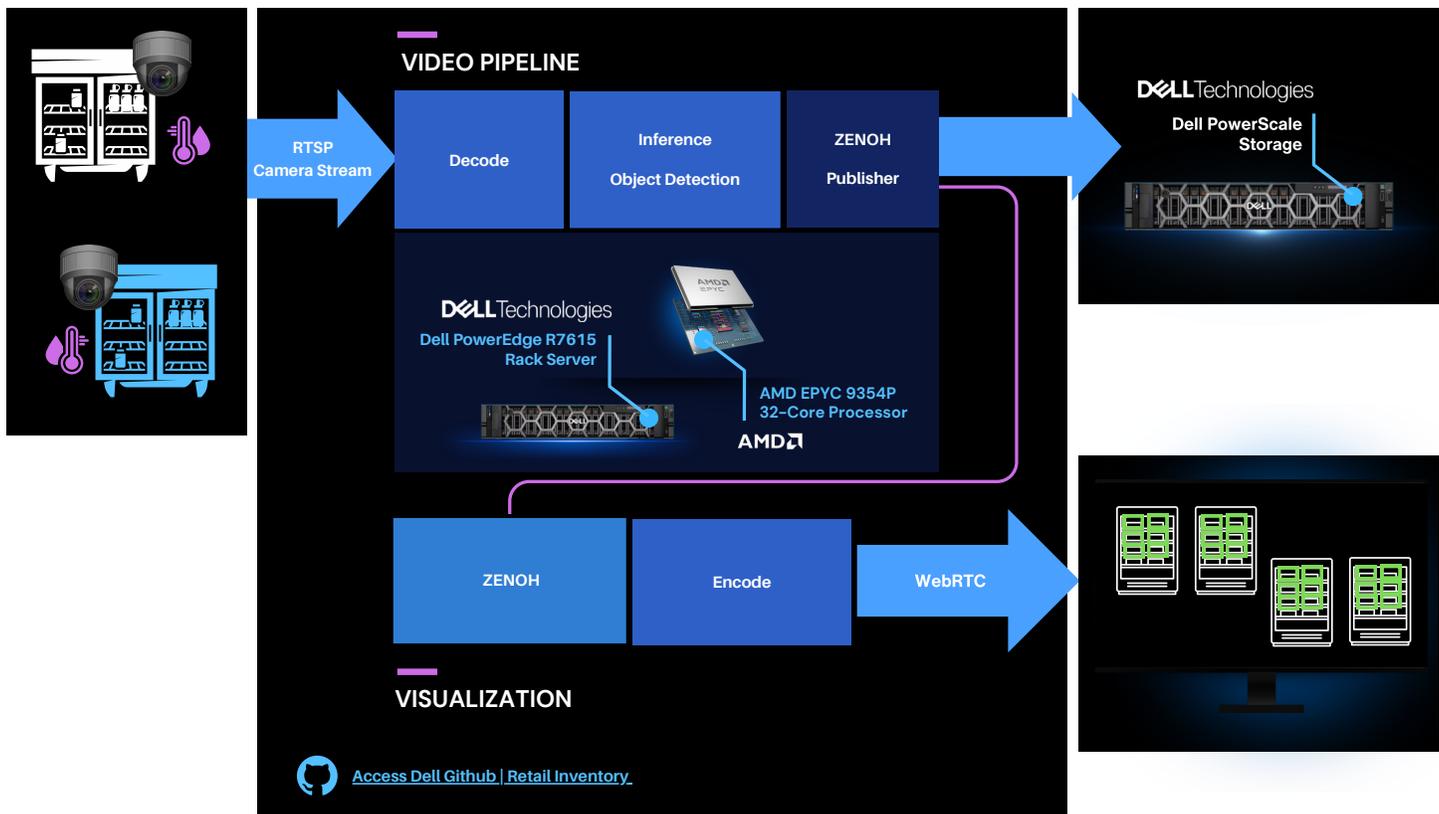


Figura 2: Diagrama de la arquitectura de la solución de referencia de gestión de inventario para comercios minoristas

## SERVICIOS DE SALUD

Las imágenes médicas impulsadas por IA son inmensamente valiosas por su capacidad para mejorar los servicios de salud al mejorar la precisión y la eficiencia del diagnóstico, y proporcionar a los profesionales de los servicios de salud información precisa sobre afecciones que pueden ser difíciles de detectar a simple vista. Al automatizar el análisis de imágenes médicas, la IA reduce el tiempo necesario para el diagnóstico, lo que permite tomar de forma más rápida decisiones sobre el tratamiento y, en última instancia, mejorar los resultados de los pacientes.

Scalers AI aprovechó las capacidades del servidor Dell PowerEdge R7625 equipado con procesadores AMD EPYC 9554 de 64 núcleos para crear una solución de imágenes médicas con tecnología de IA para la detección de neumonía. Mediante algoritmos avanzados y técnicas de aprendizaje automático para analizar imágenes médicas, como radiografías o tomografías computarizadas (TAC), la solución ayuda a aumentar la velocidad y la precisión del diagnóstico de neumonía en los pacientes. En última instancia, esto introduce una capa adicional de revisión asistida por ordenador, lo que crea un potencial para ayudar a los profesionales de los servicios de salud a gestionar grandes volúmenes de datos de imágenes de manera más eficiente.

Esta solución de referencia utiliza el modelo ResNet50 para analizar las imágenes de radiografías de tórax obtenidas del conjunto de datos del NIH Clinical Center. Su objetivo principal es detectar la presencia o ausencia de neumonía, realizando básicamente una clasificación binaria. El modelo se entrenó utilizando el conjunto de datos Xray DICOM del conjunto de datos del NIH Clinical Center, que implicó el aprendizaje por transferencia con la arquitectura ResNet50.

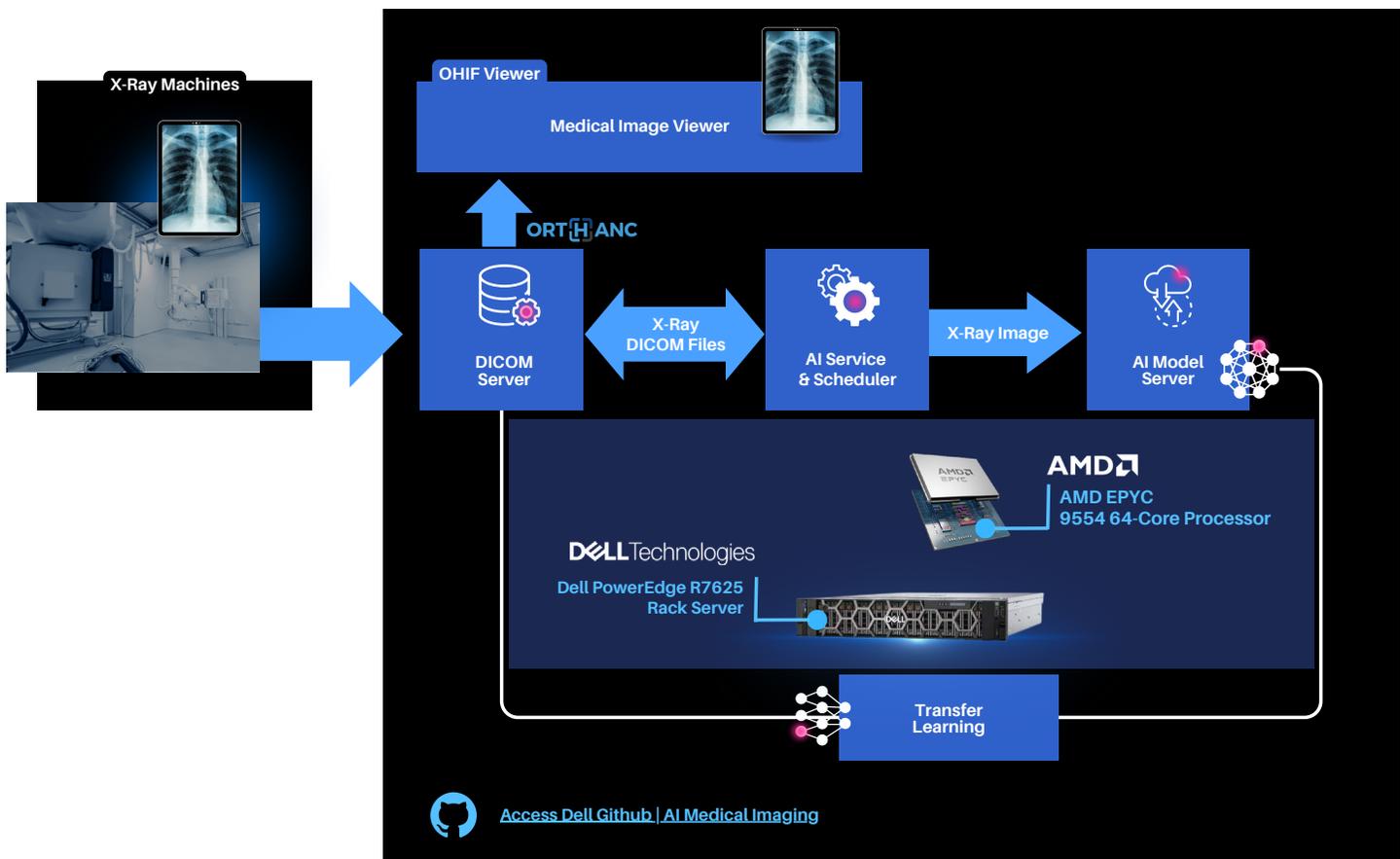


Figura 3: Diagrama de la arquitectura de la solución de imágenes médicas con IA

# Nuestra soluciones

## LA IA ES PARA TODOS: DELL Y AMD DEMOCRATIZAN LA IA

Esta colaboración sienta las bases para la democratización de la IA, que es esencial para fomentar la innovación y promover la inclusión en el ecosistema de la IA. Dell y AMD están logrando este resultado al capacitar a las personas y las organizaciones para que aprovechen la IA, y resuelvan desafíos únicos en sus respectivos campos con una suite accesible de potentes servidores equipados con las tecnologías vanguardistas de aceleradores y CPU de AMD. Los servidores Dell PowerEdge con los aceleradores AMD Instinct MI300X pueden gestionar grandes cargas de trabajo de IA, como el entrenamiento y el ajuste preciso de grandes modelos de lenguaje (LLM, por sus siglas en inglés), mientras que los servidores Dell PowerEdge equipados con procesadores AMD EPYC destacan en la gestión de cargas de trabajo de inferencia en el perímetro. Además de la plataforma de hardware subyacente, AMD también ofrece la biblioteca de software ZenDNN para la optimización de la inferencia de aprendizaje profundo en las CPU de AMD, así como la biblioteca de software AMD ROCm para mejorar el entrenamiento, el ajuste preciso y las capacidades de inferencia en los aceleradores AMD Instinct. Todas estas opciones están perfectamente integradas en el modelo Unified Inferencing Model (UIF) de AMD, a través del cual los usuarios pueden construir soluciones de IA integrales, con flexibilidad en la elección de marcos de software, optimizaciones de software y opciones de plataforma de hardware.



## COLABORACIÓN CON HUGGING FACE

Las empresas que deseen adoptar la IA pueden empezar aprovechando los modelos preexistentes o los flujos de trabajo de IA adaptados a sus necesidades específicas directamente desde Hugging Face, una plataforma de código abierto dedicada a la ciencia de datos y el aprendizaje automático. AMD ha comenzado a colaborar con Hugging Face con el objetivo compartido de ofrecer un rendimiento transformador de primer nivel mediante la adición de optimizaciones de software específicas de AMD a bibliotecas y marcos de software que ya se integran a la perfección con las plataformas AMD. De forma activa, Hugging Face está colaborando con el equipo de ingeniería de AMD para optimizar los modelos clave con el fin de obtener el máximo rendimiento, incorporando AMD ROCm en la biblioteca Transformers y mejorando Optimum-AMD, una biblioteca diseñada específicamente para plataformas AMD, con el fin de ayudar a los usuarios de Hugging Face a utilizar estos modelos con cambios mínimos en el código.

Dell Technologies también ha unido fuerzas recientemente con Hugging Face para simplificar el proceso con el que las empresas desarrollan, ajustan de forma precisa y aplican sus propios modelos de IA generativa de código abierto a través de la comunidad de Hugging Face, todo ello en relación con los productos y servicios de infraestructura de Dell líderes del sector. Se está desarrollando un nuevo portal de Dell en la plataforma Hugging Face, que incluirá scripts y contenedores personalizados y dedicados para ayudar a los usuarios a implementar de forma segura y sin esfuerzo modelos de código abierto disponibles en Hugging Face con los servidores y sistemas de almacenamiento de datos de Dell. Ahora, las empresas pueden aprovechar al máximo los recursos de Hugging Face para implementar directamente modelos en servidores Dell PowerEdge con procesadores AMD y crear soluciones de IA integrales con sus propios datos.

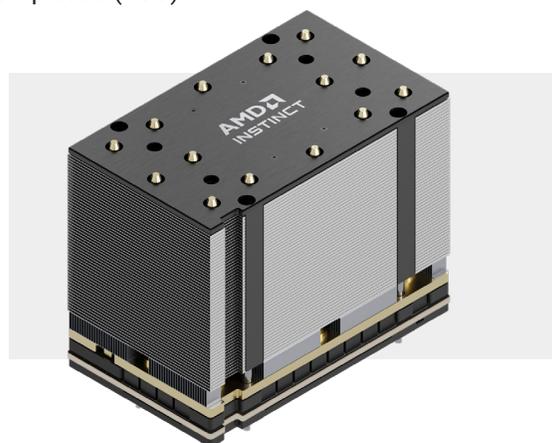


## PROCESADORES AMD EPYC

AMD ofrece los avances tecnológicos necesarios para los centros de datos modernos basados en el cloud a través de sus procesadores AMD EPYC. Estos procesadores son un sistema en chip (SoC) diseñado desde cero para satisfacer de manera eficiente las demandas de los centros de datos actuales y futuros. Los procesadores de la serie AMD EPYC 9000 equipan el centro de datos con hasta 128 núcleos, 256 subprocesos, 12 canales de memoria que admiten hasta 6 TB de memoria por socket y 128 carriles PCIe Gen5. Esta potencia se combina con la solución de seguridad de servidor x86 integrada por hardware pionera del sector. Al integrar los recursos esenciales de computación, memoria, E/S y seguridad en el SoC, los procesadores AMD EPYC ofrecen un rendimiento de nivel superior y facilitan un menor coste total de propiedad (TCO).

## ACELERADORES AMD INSTINCT MI300X

El acelerador AMD Instinct MI300X, basado en la arquitectura AMD CDNA 3 de vanguardia, ofrece eficiencia y rendimiento líderes del sector para las aplicaciones de IA y HPC más intensivas. Está equipado con 304 unidades de informática de alto rendimiento y cuenta con funciones específicas de IA, como soporte para nuevos tipos de datos y decodificación de fotos y vídeos, así como 192 GB de memoria HBM3 sin precedentes en un solo acelerador.

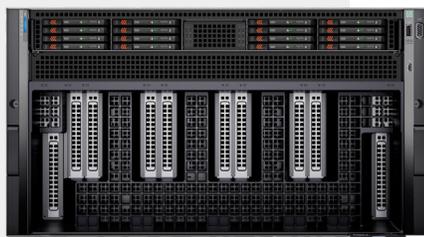


## PLATAFORMA DE SOFTWARE DE CÓDIGO ABIERTO AMD ROCm 6

La plataforma de software de código abierto AMD ROCm 6 se ha optimizado para maximizar la informática de alto rendimiento (HPC) y el rendimiento de las cargas de trabajo de IA de los aceleradores AMD Instinct MI300X. También amplía la compatibilidad con los aceleradores AMD Instinct MI300X, lo que garantiza la compatibilidad con los marcos de software del sector. La plataforma AMD ROCm encapsula una amplia variedad de controladores, herramientas de desarrollo y API que facilitan la programación del acelerador desde el nivel del kernel hasta las aplicaciones de usuario final, y se pueden adaptar para alinearse con sus requisitos específicos. AMD ROCm es especialmente adecuado para aplicaciones de informática de alto rendimiento (HPC), inteligencia artificial (IA) y computación científica. Además, la plataforma AMD ROCm ofrece compatibilidad con computación con varios aceleradores, incluido el acceso remoto a memoria directa (RDMA, por sus siglas en inglés) para la comunicación entre el servidor y el nodo.

AMD  
ROCm

## CATÁLOGO DE SERVIDORES DELL POWEREDGE



La inversión de Dell en AMD crea una opción fundamental en el mercado para democratizar la IA, como lo demuestran sus cuatro plataformas de servidor con EPYC y su servidor de montaje en rack insignia Dell PowerEdge XE9680 con aceleradores AMD Instinct MI300X. La última generación de servidores Dell PowerEdge basados en procesadores AMD EPYC mejora tanto la agilidad empresarial como el plazo de comercialización, con la capacidad de admitir cargas de trabajo transformativas, como bases de datos y análisis, virtualización, almacenamiento definido por software, infraestructura de escritorios virtuales (VDI), uso de contenedores, informática de alto rendimiento (HPC), IA y aprendizaje automático (ML). Sus servidores de montaje en rack de un socket (CPU única) proporcionan un equilibrio rentable entre rendimiento y capacidad de almacenamiento, y están diseñados para crecer a la perfección con su empresa,

mientras que sus servidores de montaje en rack de dos sockets (CPU doble) se adaptan a cargas de trabajo más exigentes con una amplia gama de funciones.

El servidor de montaje en rack Dell PowerEdge XE9680 es una sólida máquina para el procesamiento de datos diseñada especialmente para tareas de IA. Compatible con ocho aceleradores, es perfecto para cargas de trabajo de entrenamiento e inferencia de aprendizaje automático (ML) y profundo (DL), en especial las que entrenan modelos grandes de lenguaje (LLM). Equipado con ocho aceleradores MI300X, cada uno de ellos con 192 GB de memoria de gran ancho de banda de 5,3 TB/s (HBM3), lo que da lugar a una capacidad total de HBM3 de 1,5 TB por servidor y más de 21 petaflops de rendimiento FP16, el servidor de montaje en rack Dell PowerEdge XE9680 con aceleradores AMD Instinct MI300X está preparado para ampliar aún más el acceso a la IA generativa para las empresas. Esto les permite entrenar modelos más grandes, minimizar las superficies de los centros de datos, reducir el coste total de propiedad (TCO) y obtener una ventaja competitiva.

# Resumen

---

La rapidez en el ritmo de la innovación, impulsada por la IA, está revolucionando las cargas de trabajo de los centros de datos más rápido que cualquier otra transformación tecnológica. Para respaldar estos avances tecnológicos, Dell y AMD trabajan juntos para lograr un ecosistema de IA más inclusivo, innovador y desarrollado éticamente que anime a los desarrolladores de todos los sectores a colaborar en recursos de código abierto e impulsar la innovación de la IA generativa de hoy en día. Tanto si su solución de IA satisface sus requisitos de rendimiento con procesadores AMD EPYC o con servidores equipados con aceleradores AMD Instinct, le proporcionamos la flexibilidad necesaria para ejecutar sus cargas de trabajo de IA en nuestras plataformas de hardware, lo que le permite aprovechar lo mejor que Dell y AMD tienen para ofrecer.

## REFERENCIAS

Imágenes de AMD: AMD.es, Biblioteca de recursos para socios de AMD, <https://www.amd.com/en/partner/resources/resource-library.html>

Imágenes de Dell: [Dell.com](https://www.dell.com)