[notasdeprensa.jpg](http://www.mexicopress.com.mx)Publicado en Ciudad de México el 22/10/2020

# [NVIDIA Enterprise rompe récords de rendimiento en la Inferencia de IA](http://www.notasdeprensa.es)

## NVIDIA Enterprise amplía el liderazgo en MLPerf Benchmark con A100 que ofrece una inferencia de IA hasta 237 veces más rápida que las CPUs, lo que permite a las empresas trasladar la IA de la investigación a la producción

NVIDIA anunció hoy que su plataforma informática de Inteligencia Artificial ha vuelto a batir récords de rendimiento en la última ronda de pruebas de MLPerf, ampliando su liderazgo en el único benchmark (punto de referencia) independiente de la industria, que mide el rendimiento de inteligencia artificial de hardware, software y servicios. NVIDIA ganó todas las pruebas en las seis áreas de aplicación para centros de datos y sistemas informáticos en el borde en la segunda versión de MLPerf Inference. Las pruebas se expandieron más allá de las dos originales para la visión por computadora, para incluir cuatro más que cubren las áreas de más rápido crecimiento en IA: sistemas de recomendación, comprensión del lenguaje natural, reconocimiento de voz e imágenes médicas. Las organizaciones de una amplia gama de industrias ya están aprovechando el excepcional rendimiento de inferencia de la NVIDIA®A100 Tensor Core GPU and #39;s para llevar la IA de sus grupos de investigación a las operaciones diarias. Las instituciones financieras están utilizando IA conversacional para responder más rápidamente a las preguntas de los clientes; los minoristas están utilizando IA para mantener los estantes abastecidos; y los proveedores de atención médica están utilizando IA para analizar millones de imágenes médicas para identificar enfermedades con mayor precisión y ayudar a salvar vidas. "Estamos en un punto de inflexión, ya que cada industria busca mejores formas de aplicar la IA para ofrecer nuevos servicios y hacer crecer su negocio", dijo Ian Buck, Gerente General y Vicepresidente de Computación Acelerada de NVIDIA. "El trabajo que hemos realizado para lograr estos resultados en MLPerf brinda a las empresas un nuevo nivel de rendimiento de IA para mejorar nuestra vida cotidiana". Los últimos MLPerf results se producen cuando la huella de NVIDIA para la inferencia de IA ha crecido de manera espectacular. Hace cinco años, sólo un puñado de empresas líderes de alta tecnología usaban GPU’s para inferencias. Ahora, con la plataforma de inteligencia artificial de NVIDIA disponible a través de todos los principales proveedores de infraestructura de centros de datos y nube, las empresas que representan una amplia gama de industrias están utilizando su plataforma de inferencia de inteligencia artificial para mejorar sus operaciones comerciales y ofrecer servicios adicionales. Además, por primera vez, las GPUs de NVIDIA ahora ofrecen más capacidad de inferencia de IA en la nube pública que las CPUs. La capacidad informática total de inferencia de IA en la nube en las GPUs NVIDIA ha aumentado aproximadamente 10 veces cada dos años. NVIDIA lleva la inferencia de IA a nuevas AlturasNVIDIA y sus socios presentaron sus resultados de MLPerf 0.7 utilizando la plataforma de aceleración de NVIDIA, que incluye la GPU de centro de datos de NVIDIA, aceleradores de inteligencia artificial en el borde y software optimizado para NVIDIA. NVIDIA A100, presentada a principios de este año y con tecnología Tensor Cores de tercera generación y GPUs de instancias múltiples, aumentó su ventaja en la prueba ResNet-50, superando a los sistemas sólo con CPUs en 30 veces frente a 6 veces en la última ronda. Además, A100 superó a las últimas CPUs hasta en 237 veces en la prueba de recomendación recién agregada para la inferencia del centro de datos, de acuerdo con las pruebas de referencia MLPerf Inference 0.7. Esto significa que un solo servidor NVIDIA DGX A100 SYSTEM puede proveer el mismo rendimiento que alrededor de 1,000 servidores de CPUs de doble socket, ofreciendo a los clientes una rentabilidad extrema al llevar sus modelos de recomendación de IA desde la investigación hasta la producción. Los puntos de referencia también muestran que la NVIDIA T4 Tensor Core GPU sigue siendo una plataforma de inferencia sólida para empresas convencionales, servidores en el borde e instancias de nube rentables. Las GPU NVIDIA T4 superan a las CPUs hasta en 28 veces en las mismas pruebas. Además, el NVIDIA Jestson AGX Xavier™ de 30 W es el líder en rendimiento entre los dispositivos periféricos. Lograr estos resultados requirió una pila de software altamente optimizada que incluye el optimizador de inferencia el NVIDIA Triton inference serving software, ambos disponibles en NGC™, el catálogo de software de NVIDIA. Además de las presentaciones de NVIDIA, 11 socios de NVIDIA enviaron un total de 1,029 resultados utilizando las GPUs de NVIDIA, lo que representa más del 85 por ciento de las presentaciones totales en las categorías de centro de datos y en el borde.

**Datos de contacto:**

Carlos Valencia

NVIDIA PR Agency.

55 39 64 96 00

Nota de prensa publicada en: [https://www.mexicopress.com.mx/nvidia-enterprise-rompe-records-de-rendimiento](http://www.notasdeprensa.es/educalivecom-disfruta-de-cursos-presenciales-y-clases-particulares-a-traves-de-internet)

Categorías: Inteligencia Artificial y Robótica Programación Hardware Software Ciudad de México Otras Industrias

[notasdeprensa.jpg](http://www.mexicopress.com.mx)

[**http://www.mexicopress.com.mx**](http://www.notasdeprensa.es)