

Inversión y Características de una Unidad de Trabajo Inteligente para la Orquestación de Escenarios con un LLM Local: Un Análisis Comparativo de Jetson Nano y un PC

1. Introducción:

Una unidad de trabajo inteligente (UIT) se define en este contexto como una plataforma de hardware y software dedicada a la ejecución de tareas específicas, en este caso, la operación de un "orquestador de escenarios" que utiliza un modelo de lenguaje grande (LLM) de ejecución local. El orquestador de escenarios, por su parte, se concibe como un módulo encargado de la gestión y ejecución de secuencias predefinidas de acciones o simulaciones. Estas secuencias pueden implicar la interacción con otros módulos operativos y con un LLM para diversas funciones, como la generación de escenarios, la interpretación de resultados o la toma de decisiones dentro del proceso de orquestación.

La operación local de un LLM presenta diversas ventajas significativas. Entre ellas se incluyen la mejora de la privacidad de los datos, al evitar la transmisión de información sensible a servidores externos ¹, la reducción de la latencia en las respuestas, crucial para aplicaciones en tiempo real ², y la posibilidad de funcionar en entornos con conectividad a internet limitada o nula.¹

Este informe tiene como objetivo principal comparar la eficiencia de dos opciones de hardware para operar este orquestador de escenarios con un LLM local: la placa de desarrollo NVIDIA Jetson Nano y un PC de escritorio equipado con una CPU, una GPU NVIDIA dedicada, memoria RAM y una unidad de estado sólido (SSD). La finalidad última es proporcionar un análisis exhaustivo y basado en datos que sirva de guía para la toma de decisiones en la adquisición de hardware y el diseño de la arquitectura del sistema.

2. Entendiendo el Orquestador de Escenarios y sus Requisitos:

El orquestador de escenarios, como componente central de la UIT, debe poseer una serie de capacidades funcionales para cumplir con su cometido. En primer lugar, debe ser capaz de interactuar de manera fluida con un LLM de ejecución local. Esta interacción puede manifestarse en diversas formas, incluyendo la solicitud al LLM para la generación de nuevos escenarios basados en parámetros predefinidos, la interpretación de los resultados obtenidos durante la ejecución de un escenario, o incluso la delegación de decisiones específicas al LLM dentro del flujo de trabajo de la orquestación.¹

En segundo lugar, el orquestador debe gestionar de forma eficiente el ciclo de vida de los escenarios. Esto implica la capacidad de iniciar la ejecución de un escenario en respuesta a ciertos eventos o condiciones, supervisar el progreso de las acciones dentro del escenario, y

gestionar la salida o los resultados generados. Además, se anticipa que el orquestador deberá interactuar con futuros módulos operativos y de almacenamiento logístico, lo que requerirá una arquitectura flexible y extensible. Finalmente, se contempla la integración del orquestador con un Protocolo de Contexto del Modelo (MCP) para facilitar la gestión de tareas administrativas, financieras, comerciales y de marketing.¹⁶

Desde una perspectiva técnica, el orquestador de escenarios requerirá una serie de especificaciones para operar de manera eficiente ¹²²⁻¹²⁴. Esto incluye potencia de cómputo suficiente para ejecutar el LLM y la lógica de orquestación, memoria RAM adecuada para cargar el LLM, el sistema operativo y el software del orquestador ³⁵, almacenamiento SSD para el modelo LLM, conjuntos de datos y archivos de aplicación ⁴⁴, interfaces necesarias para la conexión con otros módulos y sistemas externos, y compatibilidad con el sistema operativo y los marcos de software elegidos.³ La magnitud de estos requisitos estará intrínsecamente ligada al tamaño y la complejidad del LLM seleccionado y a la naturaleza de los escenarios que se pretenden orquestar. Por lo tanto, es crucial considerar diferentes tamaños de LLM y sus correspondientes demandas de recursos.

3. Análisis Detallado de las Opciones de Hardware:

****3.1 NVIDIA Jetson Nano:****

La NVIDIA Jetson Nano se presenta como una placa de desarrollo compacta y de bajo consumo, diseñada específicamente para aplicaciones de inteligencia artificial y robótica.[2] Su arquitectura se basa en una CPU ARM Cortex-A57 de cuatro núcleos que opera a 1.43 GHz y una GPU NVIDIA Maxwell con 128 núcleos CUDA.[2, 49, 50, 51, 52, 54, 55] La placa cuenta con 4 GB de memoria LPDDR4 de 64 bits que proporciona un ancho de banda de 25.6 GB/s, y para el almacenamiento, integra una memoria eMMC 5.1 de 16 GB junto con una ranura para tarjetas microSD que permite la expansión del almacenamiento.[49, 50, 51, 52, 54, 55]

Uno de los aspectos más destacados de la Jetson Nano es su eficiencia energética, con un consumo que oscila entre 5 y 10 vatios.[2, 12, 49, 50, 51, 52, 54, 55] En cuanto al costo, el módulo base se sitúa alrededor de los \$99 para compras en volumen (1000 unidades o más), mientras que los kits de desarrollo varían en precio.[49, 51, 55, 56, 57, 58] Es importante notar que los precios en Colombia tienden a ser significativamente más altos que el precio base del módulo.[59, 60, 61, 62] Existe una versión más potente, la Jetson Orin Nano Super, que se ofrece a un precio de \$249.[63, 64, 65] En términos de conectividad, la Jetson Nano proporciona una variedad de interfaces, incluyendo puertos USB 3.0, una salida HDMI, pines GPIO y una ranura M.2 Key E.[2, 49, 50, 51, 52, 54, 55]

* ****Insight:**** El bajo consumo de energía y el factor de forma compacto de la Jetson Nano la convierten en una opción atractiva para aplicaciones embebidas. Sin embargo, su limitada memoria RAM y potencia de procesamiento podrían representar desafíos para la ejecución de modelos LLM más grandes. Además, el costo en Colombia es considerablemente superior al

precio base del módulo.

****3.2 PC con Componentes Dedicados:****

Un PC diseñado para operar un orquestador de escenarios con un LLM local típicamente incorporaría los siguientes componentes: un procesador Intel Core i5 de 12^a generación (o similar) con 6 núcleos y una frecuencia de hasta 4.40 GHz [66, 67, 68, 69] (los precios en Colombia varían [70, 71, 72]), una GPU NVIDIA dedicada, considerando una RTX 3060 con 12 GB de VRAM como una opción de gama media [73, 74, 75, 76, 77] (los precios en Colombia también varían [78, 79, 72, 80]); para LLMs más grandes, se podrían considerar opciones como la RTX 3090 con 24 GB de VRAM o tarjetas de gama superior [81, 82]), 16 GB de memoria RAM DDR4 a 3200 MHz (con capacidad de expansión) [83, 84, 85, 86] (los costos en Colombia son relativamente accesibles [87, 88, 86]), una unidad de estado sólido (SSD) NVMe PCIe 3.0/4.0 de 1 TB para una carga rápida del modelo y una operación eficiente [89, 90, 91, 92, 93, 94] (los precios en Colombia son razonables [93]), una placa base con socket LGA 1700 (compatible con Intel de 12^a generación) con suficientes ranuras PCIe y RAM [95, 96, 97, 72, 98, 99] (los costos en Colombia varían [72]), y una fuente de alimentación de 650W con certificación 80+ Bronze para asegurar el suministro energético de los componentes [100, 101, 102, 103, 104, 105] (los costos en Colombia están disponibles [102]).

En comparación con la Jetson Nano, un PC de estas características tendrá un consumo de energía considerablemente mayor y un costo inicial más elevado. Sin embargo, ofrece una mayor flexibilidad en términos de selección de componentes y capacidad de actualización, lo que podría traducirse en una mejor relación precio/rendimiento, especialmente en el largo plazo y dependiendo de las necesidades específicas.

* ****Insight:**** Un PC proporciona una flexibilidad significativamente mayor en términos de selección y actualización de componentes. El potencial para una mayor cantidad de RAM y VRAM, junto con una CPU y GPU más potentes, lo hacen más adecuado para ejecutar LLMs más grandes y manejar tareas de orquestación más complejas.

4. Análisis Comparativo de la Eficiencia:

Al comparar directamente el rendimiento computacional para la inferencia de LLMs y la orquestación de escenarios, se observa que la Jetson Nano ofrece 472 GFLOPS (FP16) 2, mientras que la Orin Nano Super alcanza 67 TOPS (107]. Estas cifras son considerablemente inferiores a las que puede ofrecer una GPU de PC dedicada como la RTX 3060, que proporciona varios TFLOPs de rendimiento FP16.106 Un PC equipado con una RTX 3060 mostrará un rendimiento significativamente mayor en términos de rendimiento y menor latencia para la inferencia de LLMs en comparación con la Jetson Nano.108 El rendimiento de la Jetson Nano para tareas de procesamiento del lenguaje natural (NLP), incluso con optimizaciones utilizando TensorRT, es limitado en comparación con las GPUs de escritorio.110

La Jetson Nano podría tener dificultades para ejecutar LLMs más grandes debido a su limitación de 4 GB de RAM ⁴², mientras que un PC con 16 GB o más de RAM puede acomodar

modelos más extensos.³⁶ Las pruebas de rendimiento en la Jetson Orin Nano Super demuestran su capacidad para ejecutar modelos de hasta 8 mil millones de parámetros, con mejoras significativas sobre la Nano original.¹⁰⁷ La Orin Nano Super alcanza velocidades de 19.14 tokens/segundo para Llama 3.1 8B y 43.07 tokens/segundo para Llama 3.2 3B.¹¹³ En cuanto a la eficiencia energética, la Jetson Nano destaca con un consumo de 5-10W, notablemente inferior al de un PC, que puede consumir varios cientos de vatios bajo carga.¹⁰⁵ Esto hace que la Jetson Nano sea adecuada para entornos con restricciones de energía.² Desde la perspectiva de la rentabilidad, la Jetson Nano tiene un costo inicial más bajo, pero el PC ofrece un mayor rendimiento por su precio, especialmente al considerar las necesidades a largo plazo y el potencial de actualizaciones.⁸¹ El costo total de propiedad, incluyendo el consumo eléctrico, podría favorecer a la Jetson Nano en escenarios muy específicos de bajo uso.

En términos de escalabilidad para futuros módulos, el PC presenta una ventaja clara debido a sus múltiples ranuras de RAM, ranuras PCIe para la adición de GPUs o almacenamiento, y una CPU más potente.⁸¹ La escalabilidad de la Jetson Nano está limitada por su configuración de hardware fija, aunque la serie Orin Nano ofrece mejores capacidades de expansión, incluyendo soporte para NVMe en la versión Super.¹⁰⁷

- **Conclusión Clave:** La configuración del PC proporciona una ventaja evidente en el rendimiento computacional necesario para ejecutar LLMs y orquestar escenarios complejos. Si bien la Jetson Nano sobresale en eficiencia energética y costo inicial, sus limitaciones en potencia de procesamiento y memoria podrían obstaculizar su capacidad para manejar LLMs más grandes y futuras expansiones. La Orin Nano Super mejora significativamente el rendimiento en comparación con la Nano original, lo que la convierte en una opción más viable para LLMs de tamaño pequeño a mediano en implementaciones embebidas.
- **Tabla 1: Comparación de Especificaciones de Hardware**

Característica	NVIDIA Jetson Nano (Original)	NVIDIA Jetson Orin Nano Super	PC (Ejemplo)
CPU	Quad-core ARM A57 @ 1.43 GHz	6-core ARM Cortex-A78AE @ 1.7 GHz	Intel Core i5 12th Gen (6 cores, hasta 4.40 GHz)
GPU	128-core NVIDIA Maxwell (472 GFLOPS FP16)	1024-core NVIDIA Ampere (67 TOPS Sparse, 17 FP16 TFLOPs)	NVIDIA RTX 3060 (12 GB VRAM)
RAM	4 GB 64-bit LPDDR4 @ 1600MHz (25.6 GB/s)	8 GB 128-bit LPDDR5 @ 102 GB/s	16 GB DDR4 @ 3200 MHz (expandible)
Almacenamiento	16 GB eMMC 5.1, microSD	Soporte para microSD y NVMe externo	1 TB NVMe PCIe 3.0/4.0
Consumo Energético	5-10 W	7-25 W	Varios cientos de W bajo carga
Conectividad	USB 3.0, HDMI, GPIO,	USB 3.2 Gen2, USB-C,	USB 3.0, HDMI,

	M.2 Key E, Gigabit Ethernet	DisplayPort, GPIO, M.2 Key M, M.2 Key E, Gigabit Ethernet	DisplayPort, Gigabit Ethernet, etc.
Precio (Colombia)	Aproximadamente COP \$1.000.000 - \$3.200.000 (Kit de Desarrollo)	No se encontraron datos específicos para Colombia	Aproximadamente COP \$1.800.000 - \$2.500.000 (sin OS)

5. Características Necesarias para el Orquestador de Escenarios:

Para que el orquestador de escenarios funcione de manera óptima, se requieren ciertas características de hardware críticas. En primer lugar, una cantidad suficiente de memoria RAM es esencial para cargar y ejecutar el LLM seleccionado.³⁶ La cantidad específica dependerá del tamaño del modelo y del nivel de cuantificación utilizado. En segundo lugar, se recomienda una GPU capaz (preferiblemente una NVIDIA RTX para aprovechar la compatibilidad con CUDA) con suficiente VRAM (al menos 6 GB, idealmente 12 GB o más para LLMs más grandes).¹⁰⁸ En tercer lugar, un almacenamiento SSD rápido (NVMe) es crucial para la carga rápida del LLM y la operación eficiente del sistema.⁴⁴ Finalmente, una conexión de red estable y confiable será necesaria para la posible integración con otros sistemas y el Protocolo de Contexto del Modelo (MCP).

Desde la perspectiva del software, el sistema operativo debe ser compatible con el hardware elegido y proporcionar los controladores necesarios. Se recomienda Linux para el desarrollo y la implementación de LLMs.¹¹⁹ Además, se requerirá soporte para la ejecución local de LLMs, ya sea a través de bibliotecas como llama.cpp, Ollama ¹²⁰, o TensorRT-LLM de NVIDIA.³ El orquestador de escenarios en sí mismo deberá ser un marco de software robusto capaz de gestionar los escenarios, interactuar con el LLM e integrarse con otros módulos. Finalmente, será imprescindible la compatibilidad con el Protocolo de Contexto del Modelo (MCP) para una integración fluida con las aplicaciones empresariales.¹⁶ La elección del LLM tendrá un impacto significativo en las características de hardware necesarias. LLMs más pequeños podrían funcionar adecuadamente en la Jetson Nano (especialmente la Orin Nano Super), mientras que los modelos más grandes y capaces requerirán los recursos que ofrece un PC. El software del orquestador de escenarios debe diseñarse con modularidad y extensibilidad para adaptarse a futuros módulos e integración con MCP.

6. Potencial para Soportar Futuros Módulos y Almacenamiento Logístico:

Al evaluar la capacidad de ambas opciones de hardware para soportar futuros módulos y almacenamiento logístico, se observan diferencias significativas. La Jetson Nano, con sus 4 GB de RAM y un único carril PCIe (la ranura M.2 Key E está destinada principalmente a conectividad inalámbrica), podría enfrentar desafíos para admitir módulos adicionales y un almacenamiento logístico considerable.² La principal vía de expansión sería a través de opciones de almacenamiento externo mediante USB. La Orin Nano ofrece un mejor ancho de banda de memoria (102 GB/s para la Super) y capacidades PCIe (ranura M.2 Key M para NVMe) ¹⁰⁷, lo que mejora sus posibilidades de expansión en comparación con la Nano original.

Por otro lado, el PC ofrece un mayor potencial de expansión.⁸¹ Un mayor número de ranuras de RAM permite aumentar la capacidad de memoria. Las múltiples ranuras PCIe pueden albergar almacenamiento adicional (SSDs NVMe o SATA) u otras tarjetas de expansión. La mayor potencia de procesamiento de la CPU puede manejar la carga de trabajo incrementada de los módulos futuros. Por lo tanto, la plataforma PC presenta una solución más escalable para soportar módulos futuros y los requisitos de procesamiento y almacenamiento asociados. La Jetson Nano podría ser adecuada para expansiones más simples, pero podría convertirse rápidamente en un cuello de botella para módulos más exigentes. La inclusión de una ranura NVMe en la Orin Nano Super mejora su capacidad de expansión de almacenamiento en comparación con la Nano original.

7. Integración con un Protocolo de Contexto del Modelo (MCP):

La integración del orquestador de escenarios con un Protocolo de Contexto del Modelo (MCP) es factible en ambas plataformas de hardware, ya que el protocolo se basa principalmente en software.¹⁶ El MCP utiliza una arquitectura cliente-servidor, donde el orquestador de escenarios (potencialmente actuando como host o cliente) se comunicaría con servidores MCP para tareas administrativas, financieras, comerciales y de marketing.²⁵ La integración implicaría la implementación de funcionalidades de cliente o servidor MCP dentro del software del orquestador, posiblemente utilizando los SDKs disponibles.¹⁷ El MCP admite varios mecanismos de transporte como STDIO, SSE y WebSockets.¹²¹ El MCP permite a los LLMs acceder a recursos (datos) y utilizar herramientas (funciones) proporcionadas por sistemas externos.²² Para el orquestador de escenarios, esto podría significar utilizar MCP para acceder a datos financieros para la planificación de escenarios, datos comerciales para el análisis de mercado dentro de los escenarios, o herramientas administrativas para la generación de informes y la gestión. El MCP simplifica la integración al proporcionar un protocolo estandarizado, reduciendo la necesidad de implementaciones personalizadas para cada herramienta o fuente de datos.¹⁸

Si bien ambas plataformas son capaces de integrarse con un MCP, la mayor potencia de procesamiento y memoria del PC podrían ser ventajosas si la integración con MCP implica un procesamiento de datos significativo o interacciones complejas con múltiples servidores. La Jetson Nano, con su menor perfil de potencia, podría ser ventajosa para una operación continua en un sistema integrado con MCP, siempre que sus recursos computacionales sean suficientes para las tareas.

8. Recomendaciones:

Considerando los factores de rendimiento, costo (incluyendo el mayor costo de la Jetson Nano en Colombia), escalabilidad y necesidades de integración, se ofrecen las siguientes recomendaciones:

- Si el enfoque principal es un consumo de energía muy bajo y un costo inicial mínimo, y el LLM a utilizar es extremadamente ligero con una orquestación de escenarios muy simple, la Jetson Nano *podría* considerarse, teniendo en cuenta sus significativas limitaciones de rendimiento y el mayor costo en Colombia.
- Para un equilibrio entre rendimiento e implementación embebida, la NVIDIA Jetson Orin Nano Super ofrece una mejora sustancial en rendimiento y características (incluyendo

soporte NVMe) con un consumo de energía aún relativamente bajo, lo que la convierte en una opción más convincente que la Nano original para ejecutar LLMs de tamaño pequeño a mediano.

- Para un rendimiento óptimo con una gama más amplia de tamaños de LLM, un soporte robusto para futuras expansiones de módulos y una integración perfecta con MCP para diversas tareas empresariales, la configuración del PC con componentes dedicados (Intel Core i5 de 12ª generación o similar, NVIDIA RTX 3060 o superior, 16 GB+ de RAM, SSD NVMe de 1 TB) es la opción fuertemente recomendada. El PC ofrece una potencia computacional, capacidad de memoria y capacidad de expansión significativamente mayores, que son cruciales para operar eficientemente los LLMs y soportar un orquestador de escenarios versátil.

9. Conclusión:

En resumen, el análisis comparativo revela que la configuración del PC ofrece un rendimiento y una escalabilidad superiores, mientras que la Jetson Nano (original) es limitada pero eficiente en cuanto a energía. La Jetson Orin Nano Super presenta un punto intermedio con un rendimiento mejorado para la IA embebida. Para una solución robusta y preparada para el futuro, capaz de manejar una variedad de LLMs e integrarse eficazmente con otros sistemas, la inversión en un PC con componentes dedicados es la opción más estratégica. La Orin Nano Super es una fuerte contendiente para implementaciones en el borde con requisitos de LLM menos exigentes. Las limitaciones de la Jetson Nano original la hacen menos adecuada para este caso de uso particular, especialmente considerando su precio en Colombia.

Obras citadas

1. Jetson Nano LLM Voice Assistant - Hackster.io, fecha de acceso: mayo 14, 2025, <https://www.hackster.io/JiongShi/jetson-nano-llm-voice-assistant-54e650>
2. Unlocking the Power of NVIDIA Jetson Nano Developer Kit for AI and Robotics - FiberMall, fecha de acceso: mayo 14, 2025, <https://www.fibermall.com/blog/nvidia-ai-board.htm>
3. Running LLMs with TensorRT-LLM on NVIDIA Jetson Orin Nano Super - Collabnix, fecha de acceso: mayo 14, 2025, <https://collabnix.com/running-llms-with-tensorrt-llm-on-nvidia-jetson-orin-nano-super/>
4. Master Jetson Nano with NVIDIA AI & GPU Computing | 21 Day Bootcamp - YouTube, fecha de acceso: mayo 14, 2025, https://www.youtube.com/watch?v=4_5vfRYJqbl
5. DeepSeek R1 on Jetson Nano: Run your LLM locally !! - YouTube, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=GdhH-CsBQvI>
6. Train an LLM From Scratch On NVIDIA Jetson Nano (Step-by-Step Guide) - YouTube, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=s1uFVfuT2aw&pp=0gcJCdgAo7VqN5tD>
7. Speech AI on NVIDIA Jetson Tutorial - JetsonHacks, fecha de acceso: mayo 14, 2025, <https://jetsonhacks.com/2023/08/07/speech-ai-on-nvidia-jetson-tutorial/>
8. dusty-nv/jetson-voice: ASR/NLP/TTS deep learning inference library for NVIDIA

- Jetson using PyTorch and TensorRT - GitHub, fecha de acceso: mayo 14, 2025, <https://github.com/dusty-nv/jetson-voice>
9. Performance Analysis: DeepSeek R1 on Jetson Nano - YouTube, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=tYTTM2Aw99M>
 10. Answering Your Questions On The NVIDIA Jetson Orin Nano Super - YouTube, fecha de acceso: mayo 14, 2025, https://www.youtube.com/watch?v=6mdwi2751y0&pp=0gcJCfcAhR29_xXO
 11. NVIDIA's \$249 Secret Weapon for Edge AI - Jetson Orin Nano Super: Driveway Monitor, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=QHBr8hekCzg>
 12. Jetson Nano Brings the Power of Modern AI to Edge Devices - NVIDIA, fecha de acceso: mayo 14, 2025, <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/product-development/>
 13. First Boot: Jetson Orin Setup with Script + LLM Help - JetsonHacks, fecha de acceso: mayo 14, 2025, <https://jetsonhacks.com/2025/04/15/first-boot-jetson-orin-setup-with-script-llm-help/>
 14. NanoVLM - Efficient Multimodal Pipeline - NVIDIA Jetson AI Lab, fecha de acceso: mayo 14, 2025, https://www.jetson-ai-lab.com/tutorial_nano-vlm.html
 15. Run Large Language Model (LLM) on NVIDIA Jetson Development Board - YouTube, fecha de acceso: mayo 14, 2025, https://www.youtube.com/watch?v=_wkltdovr6k
 16. Model Context Protocol, fecha de acceso: mayo 14, 2025, https://en.wikipedia.org/wiki/Model_Context_Protocol
 17. Model Context Protocol - GitHub, fecha de acceso: mayo 14, 2025, <https://github.com/modelcontextprotocol>
 18. Introducing the Model Context Protocol - Anthropic, fecha de acceso: mayo 14, 2025, <https://www.anthropic.com/news/model-context-protocol>
 19. Model Context Protocol (MCP) an overview - Philschmid, fecha de acceso: mayo 14, 2025, <https://www.philschmid.de/mcp-introduction>
 20. Introducing Model Context Protocol (MCP) in Copilot Studio: Simplified Integration with AI Apps and Agents - Microsoft, fecha de acceso: mayo 14, 2025, <https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/introducing-model-context-protocol-mcp-in-copilot-studio-simplified-integration-with-ai-apps-and-agents/>
 21. Specification - Model Context Protocol, fecha de acceso: mayo 14, 2025, <https://modelcontextprotocol.io/specification/2025-03-26>
 22. Model Context Protocol (MCP) Explained - Humanloop, fecha de acceso: mayo 14, 2025, <https://humanloop.com/blog/mcp>
 23. A Deep Dive Into MCP and the Future of AI Tooling | Andreessen Horowitz, fecha de acceso: mayo 14, 2025, <https://a16z.com/a-deep-dive-into-mcp-and-the-future-of-ai-tooling/>
 24. Model Context Protocol (MCP): A comprehensive introduction for developers - Stytech, fecha de acceso: mayo 14, 2025,

- <https://stytch.com/blog/model-context-protocol-introduction/>
25. What is MCP (Model Context Protocol)? - Zapier, fecha de acceso: mayo 14, 2025, <https://zapier.com/blog/mcp/>
 26. #14: What Is MCP, and Why Is Everyone – Suddenly!– Talking About It? - Hugging Face, fecha de acceso: mayo 14, 2025, <https://huggingface.co/blog/Kseniase/mcp>
 27. MCP Compatibility in AnythingLLM, fecha de acceso: mayo 14, 2025, <https://docs.anythingllm.com/mcp-compatibility/overview>
 28. Building MCP with LLMs - Model Context Protocol, fecha de acceso: mayo 14, 2025, <https://modelcontextprotocol.io/tutorials/building-mcp-with-llms>
 29. Model Context Protocol: The USB-C for AI: Simplifying LLM Integration - InfraCloud, fecha de acceso: mayo 14, 2025, <https://www.infracloud.io/blogs/model-context-protocol-simplifying-llm-integration/>
 30. Integrating Model Context Protocol Tools with Semantic Kernel: A Step-by-Step Guide, fecha de acceso: mayo 14, 2025, <https://devblogs.microsoft.com/semantic-kernel/integrating-model-context-protocol-tools-with-semantic-kernel-a-step-by-step-guide/>
 31. The Model Context Protocol (MCP): A guide for AI integration | Generative-AI - Wandb, fecha de acceso: mayo 14, 2025, <https://wandb.ai/byyoung3/Generative-AI/reports/The-Model-Context-Protocol-MCP-A-guide-for-AI-integration--VmldzoxMTgzNDgxOQ>
 32. Model Context Protocol: Introduction, fecha de acceso: mayo 14, 2025, <https://modelcontextprotocol.io/introduction>
 33. One File To Turn Any LLM into an Expert MCP Pair-Programmer : r/ClaudeAI - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/ClaudeAI/comments/1h5o9uh/one_file_to_turn_any_llm_into_an_expert_mcp/
 34. Making Sense of LLM Tool Use & MCP - YouTube, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=sVC4DL2secQ>
 35. www.rnfinity.com, fecha de acceso: mayo 14, 2025, [https://www.rnfinity.com/news-show/Hardware-requirements-for-running-large-language-model-Deepseek-R1-on-a-local-machine#:~:text=GPU%3A%20NVIDIA%20RTX%203060%20\(12.RAM%3A%2032%20GB](https://www.rnfinity.com/news-show/Hardware-requirements-for-running-large-language-model-Deepseek-R1-on-a-local-machine#:~:text=GPU%3A%20NVIDIA%20RTX%203060%20(12.RAM%3A%2032%20GB)
 36. Hardware requirements for running the large language model Deepseek R1 locally., fecha de acceso: mayo 14, 2025, <https://www.rnfinity.com/news-show/Hardware-requirements-for-running-large-language-model-Deepseek-R1-on-a-local-machine>
 37. What hardware would I need for running local LLMs? - PC - GameFAQs, fecha de acceso: mayo 14, 2025, <https://gamefaqs.gamespot.com/boards/916373-pc/80664393>
 38. How do I understand requirements to run any LLM locally? : r/LocalLLM - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/LocalLLM/comments/1hm15ox/how_do_i_understand_requirements_to_run_any_llm/
 39. How to Estimate Memory Requirements for Running Local LLMs - Msty, fecha de

- acceso: mayo 14, 2025, <https://msty.app/blog/llm-memory-requirement>
40. Using LLM in low memory (CPU) local environments - Data Science Stack Exchange, fecha de acceso: mayo 14, 2025, <https://datascience.stackexchange.com/questions/126058/using-llm-in-low-memory-cpu-local-environments>
 41. How much system RAM needed when running LLMs that can fit in GPU VRAM? - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/LocalLLaMA/comments/1e3yghg/how_much_system_ram_needed_when_running_llms_that/
 42. [D] Is it possible to run large language models using NVIDIA Jetson products? - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/MachineLearning/comments/12220vj/d_is_it_possible_to_run_large_language_models/
 43. Does the new Jetson Orin Nano Super make sense for a home setup? - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/LocalLLaMA/comments/1hh0gik/does_the_new_jetson_orin_nano_super_make_sense/
 44. Guide to Local LLMs - Scrapfly, fecha de acceso: mayo 14, 2025, <https://scrapfly.io/blog/guide-to-local-llm/>
 45. Best strategies to manage local storage while running local LLMs - SAYGE, fecha de acceso: mayo 14, 2025, <https://sayge.in/best-strategies-to-manage-local-storage-while-running-local-llms/>
 46. How to run a LLM on your machine - raresportan.com, fecha de acceso: mayo 14, 2025, <https://www.raresportan.com/how-to-run-a-llm-on-your-machine/>
 47. LLMs from external SSDs? : r/LocalLLaMA - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/LocalLLaMA/comments/1b17d8j/llms_from_external_ssds/
 48. Required Specs to Run LLMs Locally - YouTube, fecha de acceso: mayo 14, 2025, <https://www.youtube.com/watch?v=08knNUc4jgk>
 49. Jetson Nano - NVIDIA Developer, fecha de acceso: mayo 14, 2025, <https://developer.nvidia.com/embedded/jetson-nano>
 50. nvidia jetson nano developer kit | asus, fecha de acceso: mayo 14, 2025, https://dlcdnets.asus.com/pub/ASUS/mb/AIOT/AI_Accelerator/jetson-nano-devkit-datasheet.pdf
 51. NVIDIA Jetson Nano Developer Kit, Small AI Computer - WaveShare, fecha de acceso: mayo 14, 2025, <https://www.waveshare.com/jetson-nano-developer-kit.htm>
 52. NVIDIA Jetson NANO Developer Kit - SparkFun, fecha de acceso: mayo 14, 2025, <https://cdn.sparkfun.com/assets/0/7/f/9/d/jetson-nano-devkit-datasheet-updates-us-v3.pdf>
 53. How to optimize your model with TF-TRT for Jetson Nano - Marvik - Blog, fecha de acceso: mayo 14, 2025, <https://blog.marvik.ai/2021/03/12/how-to-optimize-your-model-with-tf-trt-for-jet>

[son-nano/](#)

54. NVIDIA Jetson NANO Developer Kit, fecha de acceso: mayo 14, 2025, https://dlsvr04.asus.com.cn/pub/ASUS/mb/AIOT/AI_Accelerator/jetson-nano-devkit-datasheet.pdf
55. Nvidia Jetson Nano Developer Kit - Overview - Sixfab, fecha de acceso: mayo 14, 2025, <https://sixfab.com/blog/nvidia-jetson-nano-developer-kit-a-powerful-developer-kit-for-ai-robotics/>
56. NVIDIA Jetson Nano Developer Kit (945-13450-0000-100) : Office Products - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/NVIDIA-Jetson-Nano-Developer-945-13450-0000-100/dp/B084DSDDL7>
57. NVIDIA Jetson Nano Developer Kit Detailed Review - Latest News from Seeed Studio, fecha de acceso: mayo 14, 2025, <https://www.seeedstudio.com/blog/2019/04/03/nvidia-jetson-nano-developer-kit-detailed-review/>
58. NVIDIA Jetson Nano Developer Kit - JetsonHacks, fecha de acceso: mayo 14, 2025, <https://jetsonhacks.com/2019/03/25/nvidia-jetson-nano-developer-kit/>
59. Jetson NANO 4GB Developer Kit (SUB) With Official Module For Artificial Intelligence Python Programming - RobotShop, fecha de acceso: mayo 14, 2025, <https://www.robotshop.com/products/yahboom-jetson-nano-4gb-developer-kit-sub-with-official-module-for-artificial-intelligence-python-programming>
60. Nvidia Jetson Nano Kit Para Desarrollador | MercadoLibre, fecha de acceso: mayo 14, 2025, <https://listado.mercadolibre.com.co/nvidia-jetson-nano-kit-para-desarrollador>
61. JETSON NANO DEV KIT - Sigma Electrónica, fecha de acceso: mayo 14, 2025, <https://www.sigmaelectronica.net/producto/jetson-nano-dev-kit/>
62. Nvidia Jetson Nano Development Kit (128 Cuda Cores, Colombia, fecha de acceso: mayo 14, 2025, <https://www.ubuy.com.co/en/productin/JGUM2QM3K-nvidia-jetson-nano-development-kit-128-cuda-cores-cortex-a57-4-gb-64-bit-lpddr4>
63. Jetson Orin Nano Super Developer Kit - NVIDIA, fecha de acceso: mayo 14, 2025, <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/nano-super-developer-kit/>
64. Buy the Latest Jetson Products - NVIDIA Developer, fecha de acceso: mayo 14, 2025, <https://developer.nvidia.com/buy-jetson>
65. Jetson Nano Super Development Kit announced. \$249. : r/JetsonNano - Reddit, fecha de acceso: mayo 14, 2025, https://www.reddit.com/r/JetsonNano/comments/1hgek51/jetson_nano_super_development_kit_announced_249/
66. Intel Core i5-12400F 12th Gen Alder Lake Desktop Processor, 6 Total Cores, 12 Threads, 18 MB Cache, 2.5 GHz Base Frequency, LGA 1700 Socket, 65W Power, 128 GB Max Memory Size, Box | BX8071512400FSRL4W Buy, Best Price in Oman, Muscat, Salalah - Microless Oman, fecha de acceso: mayo 14, 2025, <https://oman.microless.com/product/intel-core-i5-12400f-12th-gen-alder-lake-d>

[esktop-processor-6-core-18-mb-smart-cache-2-5-ghz-lga-1700-socket-65w-base-power-bx8071512400fsrl4w/](#)

67. Buy Intel Core I5 12400F 12th Gen Generation Desktop PC Processor CPU With 18MB Cache And Up To 4.40 GHz Clock Speed 3 Years Warranty With Fan LGA 1700 Socket (BX8071512400F) At Best Price In Siliguri, India, Kolkata, Darjeeling, Kurseong, Kalimpong, Gangtok, Sikkim, Jalpaiguri, Malbazar, Coochbehar, Malda, Guwahati, Assam, Patna, Bihar - Shivam IT Service, fecha de acceso: mayo 14, 2025, <https://shivamitservice.com/product/intel-core-i5-12400f-12th-gen-generation-desktop-pc-processor-cpu-with-18mb-cache-and-up-to-4-40-ghz-clock-speed-3-years-warranty-with-fan-lga-1700-socket/>
68. Intel Core i5-12400 - Core i5 12th Gen Alder Lake 6-Core 2.5 GHz LGA 1700 65W - eBay, fecha de acceso: mayo 14, 2025, <https://www.ebay.com/itm/166519230243>
69. Intel Core i5 (12th Gen) i5-12500 3 GHz Processor - Retail Pack - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/Intel-Core-12th-i5-12500-Processor/dp/B09MSJ29RP>
70. Procesador Intel Core I5 1235u | MercadoLibre, fecha de acceso: mayo 14, 2025, <https://listado.mercadolibre.com.co/procesador-intel-core-i5-1235u>
71. Intel Core i5-12500 12th Gen Processor - 3 GHz Colombia | Ubuy, fecha de acceso: mayo 14, 2025, <https://www.ubuy.com.co/en/product/4EVV414JY-intel-core-i5-12th-gen-i5-12500-3-ghz-processor-retail-pack>
72. Motherboard Lga 1700 | MercadoLibre - Categorías, fecha de acceso: mayo 14, 2025, <https://listado.mercadolibre.com.co/motherboard-lga-1700>
73. MSI GeForce RTX 3060 Ventus 2X 12G OC, Gaming Graphics Card - NVIDIA RTX 3060, 12GB GDDR6 Memory, 192-bit, 15 Gbps - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/MSI-GeForce-Ventus-NVIDIA-Graphics/dp/B08WHJFYM8>
74. GIGABYTE NVIDIA GeForce RTX 3060 Gaming OC V2 Graphics Card - 12GB GDDR6, 192-bit, PCI-E 4.0, 1837MHz Core Clock, RGB, 2X DP 1.4, 2X HDMI 2.1, NVIDIA Ampere - GV-N3060GAMING OC-8GD - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/Gigabyte-GeForce-GAMING-NVIDIA-GDDR6/dp/B096Y2TYKV>
75. NVIDIA Ampere Architecture for Every Gamer: GeForce RTX 3060 Available Late February, At \$329, fecha de acceso: mayo 14, 2025, <https://www.nvidia.com/en-us/geforce/news/geforce-rtx-3060/>
76. GeForce RTX 3060 Family - NVIDIA, fecha de acceso: mayo 14, 2025, <https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/rtx-3060-3060ti/>
77. GeForce RTX 30 Series Graphics Card Overview - NVIDIA, fecha de acceso: mayo 14, 2025, <https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/>
78. EVGA RTX 3060 XC GAMING - PC MASTERS BOGOTA, fecha de acceso: mayo 14, 2025,

- https://pcmastersbogota.com.co/es_es/producto/evga-rtx-3060-xc-gaming/
79. Nvidia RTX 3060 Ti: High-Performance OEM Graphics Colombia | Ubuy, fecha de acceso: mayo 14, 2025, <https://www.ubuy.com.co/en/product/ZA60K0JK-new-listing-nvidia-geforce-rtx-3060-ti-hp-oem-gpu>
80. fecha de acceso: diciembre 31, 1969, https://articulo.mercadolibre.com.co/MCO-948841948-tarjeta-de-video-nvidia-geforce-rtx-3060-ti-8gb-gddr6-_JM
81. Best Hardware for Running Large Language Models LLMs, fecha de acceso: mayo 14, 2025, <https://rational.co.in/best-hardware-for-running-large-language-models-llm/>
82. Ultimate Guide to the Best NVIDIA GPUs for Running Large Language Models, fecha de acceso: mayo 14, 2025, <https://blog.spheron.network/ultimate-guide-to-the-best-nvidia-gpus-for-running-large-language-models>
83. Memoria Ram Ddr4 16gb 2600 Mhz | MercadoLibre, fecha de acceso: mayo 14, 2025, <https://listado.mercadolibre.com.co/memoria-ram-ddr4-16gb-2600-mhz>
84. Memoria Ddr4 16gb | MercadoLibre, fecha de acceso: mayo 14, 2025, <https://listado.mercadolibre.com.co/memoria-ddr4-16gb>
85. Lexar 16GB Single DDR4 UDIMM RAM 3200MT/s CL22 288-Pin Desktop Memory, Black (LD4AU016G-B3200GSST) - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/Lexar-Single-288-Pin-Desktop-LD4AU016G-B3200GSST/dp/B08T6FW6V7>
86. Memoria RAM para portátil DDR4 PC4-25600 16GB 3200MHZ CL22 1.2V laptop, fecha de acceso: mayo 14, 2025, <https://pcsforall.co/products/memoria-ram-para-portatil-ddr4-pc4-25600-16gb-3200mhz-cl22-1-2v-laptop-marca-kingston-kvr32s22s8-16>
87. Gaming Memory| RGB DDR4 Ram - v-color Technology Inc., fecha de acceso: mayo 14, 2025, <https://v-color.net/collections/gaming-overclocking-memory>
88. Dahua 16GB RAM DDR4 3200MHz Memory Module for Laptops (DHI-DDR-C300S16G32), fecha de acceso: mayo 14, 2025, <https://www.shophive.com/dahua-16gb-ram-ddr4-3200mhz-memory-module-for-laptops-dhi-ddr-c300s16g32/>
89. SAMSUNG 9100 PRO PCIe 5.0 NVMe M.2 Internal SSD, 1TB Capacity, V-NAND TLC Type, Up to 14,700 MB/s Sequential Read & Up to 13,300 MB/s Sequential Write, TRIM & SMART Support | MZ-VAP1T0BW Buy, Best Price. Global Shipping. - Microless, fecha de acceso: mayo 14, 2025, <https://global.microless.com/product/samsung-9100-pro-pcie-5-0-nvme-m-2-internal-ssd-1tb-capacity-v-nand-tlc-type-up-to-14-700-mb-s-sequential-read-up-to-13-300-mb-s-sequential-write-trim-smart-support-mz-vap1t0bw/>
90. Unidad interna de estado sólido de 1 TB NGFF de alto rendimiento para computadora portátil SATA III 6Gb/s M2 SSD (1 TB, M.2 2280) : Electrónica - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/-/es/GAMERKING-SSD-M-2-2280-rendimiento/dp/B07PPD4MTX>

91. Amazon.com: Samsung 960 PRO 1TB SSD PCIe NVMe M.2 Internal SSD (MZ-V6P1TOBW), fecha de acceso: mayo 14, 2025, <https://www.amazon.com/Samsung-PCIe-NVMe-Internal-MZ-V6P1TOBW/dp/B01LYRCIPG>
92. Buy HikVision E3000 1TB NVMe PCIe M.2 Internal SSD - Krgkart.com, fecha de acceso: mayo 14, 2025, <https://krgkart.com/product/hikvision-e3000-1tb-nvme-pcie-m-2-internal-ssd/>
93. Disco de estado Sólido SSD PCI-E 4.0 M.2 2280 1TB NVME CRUCIAL P3 PLUS, fecha de acceso: mayo 14, 2025, <https://pcsforall.co/products/ssd-pci-e-3-0-m-2-2280-1tb-nvme-crucial-p2-ct1000p2ssd8-2-400-mb-s>
94. EVM 1TB NVME Internal SSD - Aryan Computer Solutions, fecha de acceso: mayo 14, 2025, <https://aryanonline.com/product/evm-m-2-nvme-1tb-pcie-next-gen-3d-tlc-nand-internal-ssd-evmnv-1tb-black/>
95. LGA 1700 Intel Motherboards | Newegg.com, fecha de acceso: mayo 14, 2025, <https://www.newegg.com/p/pl?Order=3&N=100007627+601394305>
96. Socket LGA 1700 Motherboards - Best Buy, fecha de acceso: mayo 14, 2025, https://www.bestbuy.com/site/searchpage.jsp?_dyncharset=UTF-8&browsedCategory=abcat0507008&id=pcat17071&iht=n&ks=960&list=y&q=processorsocketsv_facet%3DProcessor%20Socket~Socket%20LGA%201700&sc=Global&st=categoryid%24abcat0507008&type=page&usc=All%20Categories
97. LGA 1700 Motherboard - EliteHubs, fecha de acceso: mayo 14, 2025, <https://elitehubs.com/collections/lga-1700-motherboard>
98. Motherboards for Intel LGA 1700 CPUs - PC International, fecha de acceso: mayo 14, 2025, <https://pcinternational.co.za/product-category/computer-components/motherboard/intel-motherboards/motherboards-for-intel-lga-1700-cpus/>
99. Intel LGA 1700 Motherboards - Computer Orbit, fecha de acceso: mayo 14, 2025, <https://computerorbit.com/collections/intel-lga-1700-motherboards>
100. 650w power supply | Newegg.com, fecha de acceso: mayo 14, 2025, <https://www.newegg.com/p/pl?d=650w+power+supply>
101. GIGABYTE P650G 650W 80 Plus Gold Certified Power Supply (GP-P650G) - Amazon.com, fecha de acceso: mayo 14, 2025, <https://www.amazon.com/GIGABYTE-P650G-Certified-Power-Supply/dp/B0CSTCZMQ5>
102. Aigo GP650 Active Power 80PLUS BRONZE 650W Desktop pc Power Supply unit E-sports 850W maximum power 800W psu computer power supply - Microless, fecha de acceso: mayo 14, 2025, <https://uae.microless.com/product/aigo-gp650-active-power-80plus-bronze-650w-desktop-pc-power-supply-unit-e-sports-850w-maximum-power-800w-psu-computer-power-supply/>
103. Buy Cooler Master MWE 650 Bronze V2 650W, 80 Plus Bronze Certified, Non-Modular Power Supply At Best Price In Siliguri, India, Kolkata, Darjeeling, Kurseong, Kalimpong, Gangtok, Sikkim, Jalpaiguri, Malbazar, Coochbehar, Malda,

- Guwahati, Assam, Patna, Bihar - Shivam IT Service, fecha de acceso: mayo 14, 2025,
<https://shivamitservice.com/product/cooler-master-mwe-650-bronze-v2-650w-80-plus-bronze-certified-non-modular-power-supply/>
104. Best budget 650w psu! : r/buildapc - Reddit, fecha de acceso: mayo 14, 2025,
https://www.reddit.com/r/buildapc/comments/qovbqn/best_budget_650w_psu/
 105. How much would it cost to run a PC with a 650w PSU at max power for 24 hours - Reddit, fecha de acceso: mayo 14, 2025,
https://www.reddit.com/r/NoStupidQuestions/comments/5e89bx/how_much_would_it_cost_to_run_a_pc_with_a_650w/
 106. NVIDIA Jetson Nano and Jetson Xavier NX Comparison: Specifications, Benchmarking, Container Demos, and Custom Model Inference - Latest News from Seeed Studio, fecha de acceso: mayo 14, 2025,
<https://www.seeedstudio.com/blog/2020/06/04/nvidia-jetson-nano-and-jetson-xavier-nx-comparison-specifications-benchmarking-container-demos-and-custom-model-inference/>
 107. NVIDIA Jetson Orin Nano Developer Kit Gets a “Super” Boost | NVIDIA Technical Blog, fecha de acceso: mayo 14, 2025,
<https://developer.nvidia.com/blog/nvidia-jetson-orin-nano-developer-kit-gets-a-super-boost/>
 108. LM Studio Accelerates LLM Performance With NVIDIA GeForce RTX GPUs and CUDA 12.8, fecha de acceso: mayo 14, 2025,
<https://blogs.nvidia.com/blog/rtx-ai-garage-lmstudio-llamacpp-blackwell/>
 109. How to Choose the Best GPU for LLM: A Practical Guide, fecha de acceso: mayo 14, 2025,
<https://www.hyperstack.cloud/technical-resources/tutorials/how-to-choose-the-right-gpu-for-llm-a-practical-guide>
 110. Performance of different Jetson Nano models with TensorRT optimization. - ResearchGate, fecha de acceso: mayo 14, 2025,
https://www.researchgate.net/figure/Performance-of-different-Jetson-Nano-models-with-TensorRT-optimization_tbl1_371991297
 111. arijitx/jetson-nlp: NLP in NVIDIA Jetson Platform - GitHub, fecha de acceso: mayo 14, 2025, <https://github.com/arijitx/jetson-nlp>
 112. Just Super! \$249 Jetson Orin Nano Super Developer Kit - JetsonHacks, fecha de acceso: mayo 14, 2025,
<https://jetsonhacks.com/2024/12/17/jetson-orin-nano-super-developer-kit/>
 113. Benchmarks - NVIDIA Jetson AI Lab, fecha de acceso: mayo 14, 2025,
<https://www.jetson-ai-lab.com/benchmarks.html>
 114. Unleash Scalable AI Projects with the \$249 Jetson Orin Nano Super Kit, fecha de acceso: mayo 14, 2025,
<https://www.storagereview.com/news/unleash-scalable-ai-projects-with-the-249-jetson-orin-nano-super-kit>
 115. BEST hardware for running LLMs locally xpost from r/LocalLLama : r/LocalLLM - Reddit, fecha de acceso: mayo 14, 2025,
https://www.reddit.com/r/LocalLLM/comments/1it9x5w/best_hardware_for_runnin

- [g_llms_locally_xpost_from/](#)
116. BEST hardware for local LLMs : r/LocalLLaMA - Reddit, fecha de acceso: mayo 14, 2025,
https://www.reddit.com/r/LocalLLaMA/comments/1it9vkz/best_hardware_for_local_llms/
 117. Hardware Recommendations for Large Language Model Servers - Puget Systems, fecha de acceso: mayo 14, 2025,
<https://www.pugetsystems.com/solutions/ai-and-hpc-workstations/ai-large-language-models/hardware-recommendations/>
 118. How to Run an LLM Locally with Pieces, fecha de acceso: mayo 14, 2025,
<https://pieces.app/blog/how-to-run-an-llm-locally-with-pieces>
 119. Tech Primer: What hardware do you need to run a local LLM? | Puget Systems, fecha de acceso: mayo 14, 2025,
<https://www.pugetsystems.com/labs/articles/tech-primer-what-hardware-do-you-need-to-run-a-local-llm/>
 120. How Fast Does the Jetson Nano Really Run Large Language Models?, fecha de acceso: mayo 14, 2025,
<https://www.jeremymorgan.com/blog/tech/nvidia-jetson-orin-nano-speed-test/>
 121. Unleashing the Power of Model Context Protocol (MCP): A Game-Changer in AI Integration, fecha de acceso: mayo 14, 2025,
<https://techcommunity.microsoft.com/blog/educatordeveloperblog/unleashing-the-power-of-model-context-protocol-mcp-a-game-changer-in-ai-integration/4397564>
 122. fecha de acceso: diciembre 31, 1969,
<https://www.assemblyai.com/blog/can-the-nvidia-jetson-nano-run-modern-ai-models/>
 123. karpathy/nanoGPT: The simplest, fastest repository for ... - GitHub, fecha de acceso: mayo 14, 2025, <https://github.com/karpathy/nanoGPT>
 124. fecha de acceso: diciembre 31, 1969,
<https://developer.nvidia.com/blog/accelerating-large-language-model-inference-with-tensorrt/>