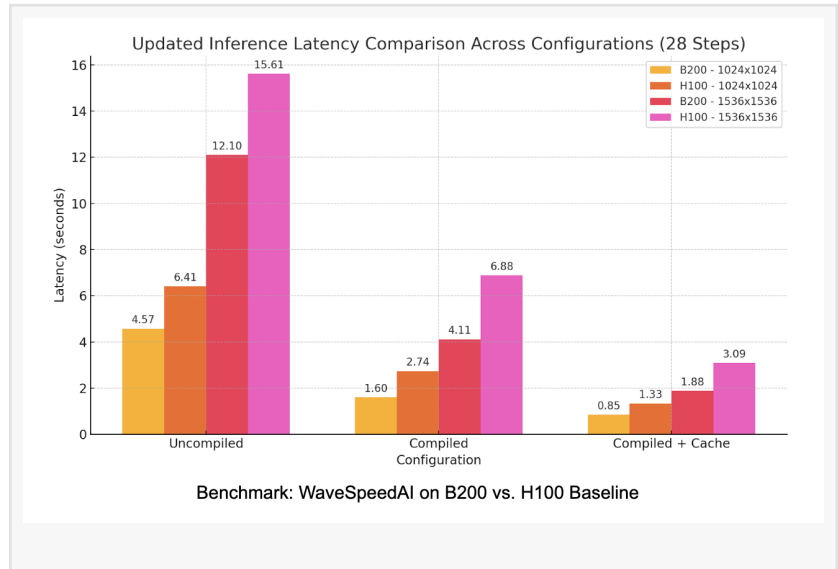# WaveSpeedAI Secures Multi-Million Dollar Funding to Build the Fastest Infrastructure for AI Image and Video Generation

WaveSpeedAI secures multimillion-dollar funding to optimize its proprietary inference engine and accelerate next-gen multimodal AI infrastructure.

SINGAPORE, SINGAPORE, April 24, 2025 /EINPresswire.com/ -- About WaveSpeedAI

WaveSpeedAI is building a next-generation infrastructure for generative AI, with a focus on high-performance inference across image, video, and audio modalities. Designed for developers, creators, and

Benchmark: WaveSpeedAI on B200 vs. H100 Baseline

businesses, the platform provides lightning-fast inference APIs and seamless ComfyUI integrations—enabling effortless deployment of complex multimodal generation tasks.

Currently, WaveSpeedAI supports a broad range of state-of-the-art models, including but not limited to:
- Image Generation & Transformation: Flux Series, Ghibli, HiDream and more
- Video Generation: Wan 2.1, Kling V1.6, Minimax and more
- Audio Generation: MMAudio v2 and more

All Flux and Wan models support full LoRA compatibility and dynamic ControlNet loading, offering maximum flexibility and customization in creative pipelines.

Proprietary Inference Framework: Up to 6x Faster, One-Third the Cost
WaveSpeedAI's technical edge lies in its proprietary dynamic compute scheduling and fused inference architecture, which extracts maximum performance across GPU platforms including A100, H100, and the latest B200 chips.

According to recent independent benchmarks ([source]), WaveSpeedAI delivers:
- Up to 6x faster inference on image transformation tasks (e.g. Flux series)

- One-third the compute cost of traditional cloud solutions
- Real-time frame generation for video models with LoRA and hardware optimization

WaveSpeedAI's universal inference engine supports both open-source and proprietary models, breaking down deployment barriers and drastically reducing time-to-market for AI-native applications.

Founding Team: Deep Technical DNA

WaveSpeedAI was founded by two seasoned engineers with extensive experience in AI infrastructure and distributed systems. The company is headquartered in Singapore, strategically positioned to serve a global developer base with low-latency access across North America, Europe and Asia-Pacific.

- Zeyi Cheng, Founder & CEO — Creator of several high-performance open-source AI optimization projects such as stable-fast, ParaAttention, and Comfy-WaveSpeed. Previously led inference engine development at a top Silicon Valley AI company, delivering a 10x performance increase that helped fuel commercial success.

- David (Yangbing) Li, Co-founder & CTO — Former lead engineer at Meicai and DP Technology, where he built large-scale distributed scheduling systems powering millions of AI tasks. A frequent speaker at events like Alibaba Cloud's Apsara Conference, he brings expertise in scalable systems and heterogeneous compute acceleration.

Together, the duo combines best-in-class algorithmic innovation with system-level engineering, giving WaveSpeedAI a rare dual advantage in this fast-evolving space.

Vision: Infrastructure for the Generative AI Era

As generative AI applications explode in scale and complexity, performance, latency, and cost have become the core bottlenecks.

WaveSpeedAI is on a mission to become the default infrastructure provider for multimodal AI generation, providing a unified inference runtime, composable model endpoints, and dynamic optimization tools. From creative content and AI agents to gaming, video, and audio applications, the platform is designed to empower developers, creators and businesses to build the next wave of AI-native products with unmatched speed and efficiency.

Serving thousands of developers globally via APIs and private deployments, WaveSpeedAI will continue to launch new cutting-edge models and performance upgrades at an accelerated pace throughout 2025.

Shunxin Pang
HashMatrix
email us here
Visit us on social media:

This press release can be viewed online at: https://www.einpresswire.com/article/806289061

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.