



Nvidia's 90% Cost Cut: Why AI Crypto Tokens are Skyrocketing This Week

Summary

At the CES 2026 tech show in Las Vegas this week, Nvidia CEO Jensen Huang unveiled the "**Vera Rubin**" platform, a new generation of AI chips that will change the economy of the internet. The biggest news is that this new hardware can reduce the cost of generating "**AI Tokens**" by a staggering 90% compared to previous models. This means the "brain power" needed to run AI agents is becoming ten times cheaper almost overnight. As a result, investors are flooding into AI-related crypto tokens, betting that cheap computing will lead to a massive explosion in the number of active AI bots worldwide.

What It Means

Nvidia's new Vera Rubin architecture isn't just a faster chip; it is a complete redesign of how AI "thinks." By combining six different types of chips into one "**Supercomputer**" rack, Nvidia has removed the bottlenecks that used to make AI slow and expensive. This "**Extreme Co-design**" allows bots to process massive amounts of data—like long conversations or complex market trades—using much less electricity and hardware.

For the crypto market, this is the "**spark**" that everyone was waiting for. Most AI crypto projects (like Render, Near, or Bittensor) rely on expensive computing power to function.

When the cost of that power drops by 90%, these projects suddenly become much more profitable and useful. If it costs 10 cents to run a bot instead of 1 dollar, millions of new users will start using AI agents for daily tasks.

This news has triggered what traders call a "**Flywheel Effect**." As the cost per token drops, the demand for these tokens increases because more people can afford to use them. In 2026, we are moving from the "**Training Phase**" (teaching AI) to the "**Inference Phase**" (actually using AI). Nvidia is providing the cheap fuel that allows this new "**Agentic Economy**" to finally take off for regular people.



Key Takeaways

- **90% Cost Reduction:** The Vera Rubin platform cuts the cost of AI reasoning by 10x
- **Vera CPU & Rubin GPU:** A new "Superchip" combo designed specifically for AI agents
- **10x Throughput:** Bots can now "think" and "talk" ten times faster than last year
- **AI Token Surge:** Crypto coins linked to AI computing saw double-digit gains this week

- **Agentic Focus:** Nvidia's new hardware is built specifically to support autonomous bots
- **Lower Entry Barrier:** Small startups can now run powerful AI models once limited to Big Tech
- **HBM4 Memory:** Ultra-fast memory allows bots to remember much longer conversations
- **Mass Adoption:** Cheaper tokens mean AI services will soon be free or very low-cost.

Our Take (Outlook 2026) * Speculative

The **Vera Rubin launch** is the most important event for AI crypto in 2026 so far. By the end of this year, the cost of running AI agents is expected to fall so low that "**AI salaries**" could become a common secondary income stream. The 90% cost cut acts as a global **green light** for the bot economy. For holders of AI-linked crypto assets, this confirms that the underlying technology is finally ready for mass-market use.

References

- Nvidia Newsroom: "NVIDIA Kicks Off Next Generation of AI With Rubin" (Jan 5, 2026)
- FinancialContent: "NVIDIA Unveils Vera Rubin AI Platform at CES 2026" (Jan 9, 2026)
- Tom's Hardware: "Nvidia launches Vera Rubin NVL72 – 10x lower cost per token" (Jan 5, 2026)
- SiliconANGLE: "Nvidia resets the economics of AI factories" (Jan 10, 2026)

CryptxAI publishes simplified AI and crypto briefings.