Kneron Enables Secure, Local Enterprise Agentic AI Through OpenClaw Integration on KNEO 350

Integration Combines Real-Time NPU Inference, Private On-Prem AI, and Local Agentic Workflows for Enterprises

Kneron Kneo 350

SAN DIEGO, CA, UNITED STATES, June 2, 2026 /EINPresswire.com/ -- Kneron, a semiconductor company delivering real time inference through energy-efficient edge AI and advanced neural processing systems, today announced growing enterprise momentum around its integration of OpenClaw on the KNEO 350 edge AI platform, enabling organizations to deploy private, agentic AI workflows locally through high-performance NPU-powered inference.

Running OpenClaw locally on KNEO 350 allows enterprises to keep sensitive workflows, proprietary knowledge, and internal business data inside their own infrastructure rather than transmitting information through external cloud AI services. Combined with Kneron's energy-efficient NPU architecture, organizations can achieve lower operating costs, reduced cooling requirements, improved compliance posture, and greater long-term control over enterprise AI deployment.

> The industry response to KNEO 350's integration with OpenClaw has been extremely strong,"
>
> *Dr. Albert Liu*

As enterprises confront growing AI infrastructure costs, KNEO 350 is emerging as a radically more deployment-efficient approach to enterprise inference. In internal benchmarking against Oracle H100 cloud infrastructure, KNEO 350 achieved near-H100-class decode throughput for GPT-OSS 20B workloads while operating within a significantly lower power envelope. For GPT-OSS 120B benchmarks, Oracle's published configuration required two H100 GPUs, highlighting the growing infrastructure demands

associated with large-scale GPU inference deployments. KNEO 350 was designed specifically for this next era of AI deployment: localized, energy-efficient, enterprise-scale inference built for real-world operational environments.

Unlike on premise NPU-based architectures, hybrid/cloud environments bring with them GPU-infrastructure dependencies, non-deterministic model behavior, and fragmented data, which make automation difficult, increase costs, and create significant security and compliance gaps.

KNEO 350 runs private AI services and local agent capabilities inside the enterprise. OpenClaw provides a more natural conversational entry point, allowing employees to interact with local agents through a familiar chat-style like experience, using tools and apps that feel like Slack, WhatsApp, or Line. The combined solution offers a simplified workflow that allows employees to ask, assign, search and interact in the most natural way possible.

“The industry response to KNEO 350’s integration with OpenClaw has been extremely strong,” said Dr. Albert Liu, founder and CEO of Kneron. “Our combination leverages the power of KNEO 350 and OpenClaw’s localized agentic AI apps and tools that are designed to maximize adoption, ease of operations and compliance – all while protecting the security and integrity of an organization’s existing infrastructure.” As one of the earliest pioneers in NPU technology, Kneron continues to evolve beyond inference silicon into full-stack edge AI infrastructure, and KNEO 350’s support for OpenClaw further reinforces our commitment to enabling private, localized, enterprise-scale AI deployment.

“Our goal is to transform the enterprise by deploying agentic AI with speed, consistency and low risk, which is why it just made sense to pair OpenClaw with KNEO 350, with its ability to support a variety of workflows and deliver NPU-powered real time inference,” added Liu.

After evaluating several AI chip solutions, Kneron’s KNEO 350 was the obvious choice for the OpenClaw project due to its agility, powerful inference engine, and its technology direction. KNEO 350 and OpenClaw together create tremendous value by reshaping the repetitive, fragmented work that enterprise teams deal with every day.

Based on the success of this collaboration, and KNEO 350’s flexibility, Kneron plans to support the deployment of other AI agent providers such as Pico-Claw or Hermas agent.

More About Kneron’s KNEO350

KNEO 350 is a high-performance edge server for the heavy AI workloads and AI acceleration that today’s industrial enterprises demand. Offering AI processing power of 1600 Trillions of Operations per Second (TOPS) and high-throughput analytics, KNEO 350 comes with pre-loaded applications like KneoChat and Kneron’s AI Agent development framework for rapid custom workflow creation.

KNEO 350 is part of Kneron's edge AI server family, designed to bring private, on-device Generative AI capabilities to the edge computing environment. The KNEO 350 series, along with the KNEO 330, represents Kneron's latest advancement in edge AI computing technology. These devices are designed to enable real-time video processing and Gen-AI applications without requiring cloud connectivity, emphasizing on-device computation for enhanced privacy.

More About Kneron Full-Stack Edge AI: From Device to Data Center

As generative AI transforms today's industries, enterprises are no longer asking if they should adopt AI -- but how to deploy it securely and sustainably. Kneron answers that question with KneoEdge™, a suite of integrated Edge AI systems engineered for everything from pilot projects to enterprise-wide rollouts.

More than just hardware, KneoEdge™ is an end-to-end AI infrastructure stack built on Kneron's proprietary NPU architecture. It seamlessly combines NPU processors, a secure operating system, an optimized inference engine, and a unified management platform—all guided by one principle: Data stays local. Intelligence runs at the edge. Management stays centralized.

About Kneron

Founded in October 2015 and headquartered in San Diego, California, Kneron is a pioneering force in Neural Processing Unit (NPU) technology and full stack edge AI, having focused on inference-optimized architectures more than a decade ago. The company's mission is to democratize AI, creating a future where every device is independently intelligent, every application is private, and every user can harness AI safely and effortlessly. By combining proprietary AI hardware, powerful software, and an open development ecosystem, Kneron empowers innovators, enterprises, and consumers to build, deploy, and scale intelligent solutions like never before.

With more than 200 employees globally and nearly $200 million in funding, backed by leading global investors including Horizons Ventures, Alibaba Entrepreneurs Fund, Qualcomm, Sequoia Capital, CDIB, Foxconn, and Himax Technologies, Kneron is driving the next generation of edge intelligence.

Andrea Corry
TopMind PR and Marketing for Kneron
+1 925-640-5482
andrea@topmindpr.com
Visit us on social media:
LinkedIn
Instagram
Facebook