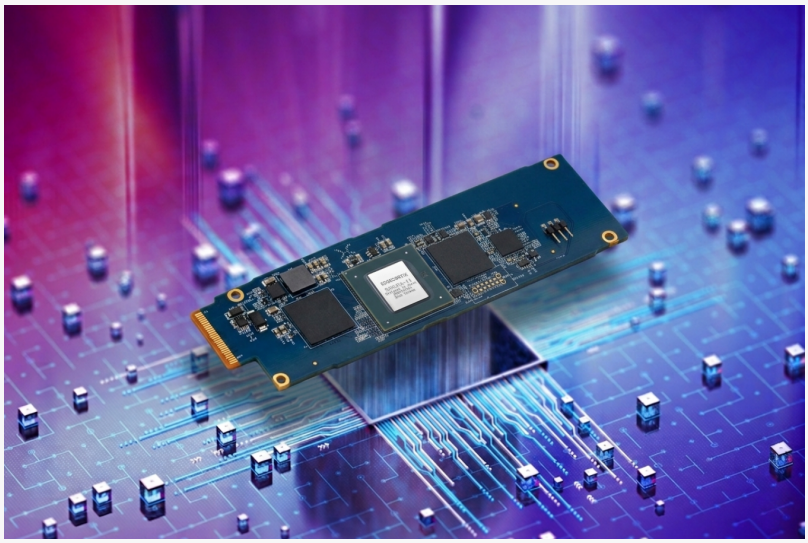
Unigen Expands AI Portfolio with High-Performance On-Prem GenAI Module

Powered by EdgeCortix SAKURA-II, Amaretti delivers 60 TOPS at just 10W, enabling secure, on-prem LLM and VLM deployment



NEWARK, CA, UNITED STATES, April 13, 2026 /EINPresswire.com/ -- Unigen Corporation, a global leader in the design and manufacturing of enterprise and industrial electronics, today announced the expansion of its AI product portfolio with the launch of the Amaretti E1.S AI Module. This release marks Unigen's first Generative AI (GenAI) hardware solution, building upon the success of its established object detection portfolio, which includes the Cupcake Edge AI Server, Biscotti E1.S AI module, Poptart E3.S AI module, and Poundcake VMS Server.

Designed to meet the skyrocketing demand for localized generative AI, Amaretti provides a high-density, low-power, on-premise solution for complex workloads. When paired with AMD or Intel servers, Amaretti is the ideal engine for Generative AI and Large Language Model (LLM) applications, supporting models with up to 20 billion parameters.

Ultra-Efficient Intelligence

The Amaretti E1.S AI Module delivers 60 trillion operations per second (TOPS) of AI computing power while consuming only 10 watts. Amaretti E1.S is powered by the EdgeCortix SAKURA-II accelerator, designed to deliver fast, real-time AI inference with high performance in a compact, low power form factor. SAKURA-II is designed to handle the most challenging GenAI applications at the edge, enabling developers to generate new content based on diverse inputs such as images, text, and audio. Amaretti E1.S features up to 32GB memory, and the module delivers an industry-leading 6 TOPS per watt. This level of efficiency allows enterprises to run up to 20 billion parameter LLMs and VLMs locally, without compromising thermal management or system power constraints.

Accelerating Your AI Market Opportunity

Amaretti E1.S enables the delivery of specialized AI hardware that maintains full compatibility with broader hardware ecosystems, offering a significant time-to-market advantage through pre-validated processing capabilities. Beyond raw text processing, Amaretti delivers low-latency performance for Text-to-Speech, Visual Language Models, and YOLO object detection. By leveraging its proprietary MERA SDK, Amaretti ensures that AI tasks never compete for resources, excelling in time-sensitive requirements for mission-critical robotics, secure AI agents, aerospace, defense and intelligent industrial monitoring applications.

Comprehensive Safe AI Framework Ecosystem

The Amaretti E1.S provides flexibility for AI architects and system integrators, featuring seamless integration with industry-standard frameworks including TensorFlow, PyTorch, ONNX, and Hugging Face. This modular approach protects infrastructure investments while delivering up to a 10x reduction in operational costs compared to cloud-dependent architectures. Crucially, Amaretti eliminates security and data privacy concerns through its on-prem architecture, allowing enterprises to leverage their proprietary IP without ever exposing sensitive data to the public web.

"As part of our expanding roadmap for Edge AI modules and servers, the Amaretti E1.S is the first in a line of modules designed to support the latest in Generative AI," said Paul W. Heng, Founder and CEO of Unigen. "By focusing on accessible, high-performance solutions for small and medium businesses, Unigen is once again leading the enterprise technology market with high-density, high-value computing."

“We’re proud to partner with Unigen and bring SAKURA-II into the Amaretti platform, enabling our energy-efficient AI to scale into server-class deployments through a high-density, modular form factor,” said Sakyasingha Dasgupta, Founder and CEO of EdgeCortix. “This is a key step in expanding the thick edge, bringing powerful and efficient AI closer to where real-world decisions happen. Together, we are not only enabling a new class of on-device systems, but also paving the way for agentic AI that can operate autonomously in dynamic environments. We’re continuing to push our silicon and software forward to support this shift toward faster, more adaptive, and energy-efficient intelligence.”

Join Unigen at the MSP Summit at The Venetian, Las Vegas (booth #MSP16), from April 13-14 to explore our AI capabilities and 2026 product roadmap. You can also find EdgeCortix at NexTech Tokyo Spring 2026 (Booth #1-30, West Hall 1) from April 15-17, where they will showcase live demonstrations of scalable edge AI and partner applications.

About Unigen Corporation

Founded in 1991, Unigen is an established global leader in the design and manufacture of OEM products including SSDs, DRAM modules, NVDIMMs, Enterprise IO and AI solutions. Unigen also offers a full array of Electronics Manufacturing Services (EMS), including design, quick-turn prototyping, new product introduction, volume production, supply chain management, assembly

& test, and aftermarket services. Headquartered in Newark, California, the company operates state-of-the-art manufacturing facilities (ISO-9001/14001/13485 and IATF 16949) in the heart of Silicon Valley as well as offshore in Vietnam and Malaysia. Unigen offers its products and services to customers worldwide targeting a broad range of end markets including automotive, computing and storage, embedded, medical, AI, robotics, clean energy, and IoT. Learn more about Unigen's products and services at Unigen.com.

Jeff Chang
Unigen Corporation
email us here
Visit us on social media:
LinkedIn
