

KALRAY DEBUTS NGENEA® FOR AI, A DATA ACCELERATION PLATFORM FOR GENAI AND RAG

Enables unmatched performance, efficiency by speeding up ingest performance and access to unstructured data from a unified, global namespace

Grenoble, France, and Las Vegas NV - May 21, 2024 – Kalray (Euronext Growth Paris: ALKAL), a technology innovator building software and hardware solutions that accelerate data-intensive workflows, announced Ngenea® for AI, a new edition of its Kalray Data Acceleration Platform specifically fine-tuned for AI data pipelines. As AI demands new ways to ingest and access massive volumes of data, Ngenea for AI helps teams speed up performance, simplify data management and open their data assets to new AI-powered use cases.

Ngenea for AI is the companion to two Ngenea editions for Media & Entertainment and HPC customers, respectively. The new Ngenea for AI empowers AI innovators to speed up their ingest performance and access their unstructured data from a unified, global name space. Kalray debuted Ngenea for AI at Dell Technologies World 2024 in Las Vegas today, where the software's performance on Dell PowerEdge servers is being showcased in the Kalray Booth #1026.

The generative AI (GenAI) market is predicted to see 42% compound annual growth rate (CAGR) across the next 10 years into a \$1.3 trillion space, from just \$40 billion in 2022 (Bloomberg Intelligence). Yet innovative AI constructs such as large language models (LLMs) and Retrieval-Augmented Generation (RAG), and applications including Smart Vision and GenAI, are challenging the ability of IT systems to move data to and from their massive data stores.

ENABLING GENAI AND RAG

Ngenea for AI adds data indexing and search capabilities, which users can leverage to feed any data to Smart Vision, GenAI and RAG applications effortlessly. The new offering enables AI customers to fully harness their existing assets when deploying GenAI applications, regardless of where the data is stored. These GenAI and RAG capabilities simplify and automate what would otherwise be a complex workflow from a data management aspect.

"Fast AI processing is mission-critical to our business of detecting railcar anomalies before they lead to breakdowns or derailments. With Ngenea deployed on Dell Technologies products, Duos achieved a 120x performance boost on our Railcar Inspection Portal (RIP®) solution, and an 8x increase in inspection accuracy. Working with Kalray helped us put railroads at the forefront of innovation," said Jeff Necciai, chief technology officer of Duos Technologies.



HIGH-PERFORMANCE STORAGE TIER TO FEED GPUS EFFORTLESSLY

The Ngenea for AI platform also incorporates a high-performance storage tier for the most data-intensive AI workloads. The tier is powered by a proven high-performance parallel file system that can easily manage petabytes of data and billions of files – whether data live on the edge, in the cloud, or on premises. This allows customers and partners to maximize the value of their GPU investment, without rip-and-replace deployments.

With Ngenea, AI innovators:

- **Accelerate AI workflows** by optimizing AI data pipelines and enhancing the overall speed and efficiency of AI operations.
- **Boost the AI model performance** by feeding data to GPUs faster, allowing for quicker computations and accelerated insights.
- **Gain advanced data indexing capabilities** that enable Generative AI and RAG applications to deliver more refined search functionalities, enabling more effective queries and a more strategic utilization of AI models.
- **Facilitate a controlled and optimized connection** between company data and a suite of third-party AI-powered tools, ensuring seamless integration and data synergy.

Optional hardware acceleration via Kalray's DPUs including the TURBOCARD4 (TC4) allows parallel processing in an asynchronous way for ultra-demanding workflows, via a complementary architecture to GPUs.

"Even with the tremendous commitment that C-suite leaders are making to AI, very few CIOs have the sizable budgets required to deploy a new system solely for an AI rollout," said Scott Sinclair, practice director, infrastructure, and modernization, at the Enterprise Strategy Group™. "Innovative use cases that enable enterprises to leverage their existing infrastructure and still deploy novel AI applications address a sizable untapped need in the market."

AI TOOLKIT FOR EASY INTEGRATION

Ngenea for AI includes an AI Toolkit that Kalray partners can use for easy integration of their applications with Ngenea. For example, when integrating Ngenea with Perifery Vision AI, users can intuitively search for historical content across a vast amount of data on locations, types, and collections.

"Traditional data centers weren't built for massive data sets found in AI workloads – or to process these workloads rapidly and efficiently," said Eric Baissus, CEO of Kalray. "Kalray's Ngenea for AI is the data acceleration platform for the AI age – software that elevates performance on existing servers and storage and provides toolkits that streamline application integration. With our software, enterprises and service providers can feed GPUs effortlessly - optimizing data movement and orchestration as the foundation of their infrastructure - so that AI genuinely delivers a competitive edge."

The Ngenea for AI platform joins two other Ngenea editions that streamline application integration for specific use cases: Ngenea for HPC – providing scratch storage, or storage for temporary user data, and Ngenea for Media & Entertainment – to accelerate post-production workflows.

Ngenea for AI is available today from Kalray and select partners. For more information visit [here](#)

ABOUT KALRAY

Kalray is a technology innovator specializing in software and hardware solutions to accelerate data-intensive workflows in Media & Entertainment, High-Performance Computing, and Artificial Intelligence. Our comprehensive product range features Ngenea, a leading data acceleration platform, as well as accelerated data processing cards for storage and compute.

With Kalray solutions, customers can scale their infrastructures to efficiently meet the performance and capacity needs of data-intensive tasks, all without being tied to specific data formats or vendors.

Founded in 2008 as a spin-off from the French CEA research lab, Kalray has corporate and financial backing from notable investors such as Alliance Venture (Renault-Nissan-Mitsubishi), NXP Semiconductors, and Bpifrance. Committed to innovation and excellence, we strive to deliver value to our customers, developers, and the planet through technology and expertise.

INVESTOR CONTACTS

Eric BAISSUS

contactinvestisseurs@kalrayinc.com

Phone +33 4 76 18 90 71

ACTUS Finance & Communication

Anne-Pauline PETUREAUX

kalray@actus.fr

Phone + 33 1 53 67 36 72

PRESS CONTACTS

Penta STANLEY

communication@kalrayinc.com

Phone +44 7939 877 880

Mary Kae Marinac

PR Representative for Kalray

mk@mkmarinac.com

Phone +1 978-685-3136

ACTUS Finance & Communication

Serena BONI

sboni@actus.fr

Phone +33 4 72 18 04 92