KIOXIA

Press Release

KIOXIA AiSAQ™ Technology Designed to Reduce DRAM Requirements in Generative AI Systems Released as Open Source Software

Software technology improves vector database scaling and accuracy in RAG workflows by using SSDs.

Germany, Düsseldorf, 29 January 2025 – KIOXIA, today announced the open-source release of its new All-in-Storage ANNS with Product Quantization (AiSAQ™) technology. A novel "approximate nearest neighbor" search (ANNS) algorithm optimized for SSDs, KIOXIA AiSAQ™[1] software delivers scalable performance for retrieval-augmented generation (RAG) without placing index data in DRAM - and instead searching directly on SSDs.

Generative AI systems demand significant compute, memory, and storage resources. While they have the potential to drive transformative breakthroughs across various industries, their deployment often comes with high costs. RAG is a critical phase of AI that refines large language models (LLMs) with data specific to the company or application.

A central component of RAG is a vector database that accumulates and converts specific data into feature vectors in the database. RAG also utilizes an ANNS algorithm, which identifies vectors that improve the model based on similarity between the accumulated and target vectors. For RAG to be effective, it must rapidly retrieve the information most relevant to a query.

Traditionally, ANNS algorithms are deployed in DRAM to achieve the high-speed performance required for these searches.

KIOXIA

KIOXIA AiSAQ™ technology provides a scalable and efficient ANNS solution for billion-scale datasets with negligible memory usage and fast index switching capabilities.

Key Benefits of KIOXIA AiSAQ™ technology:

- Allows large-scale databases to operate without relying on limited DRAM resources, enhancing the performance of RAG systems.
- Eliminates the need to load index data into DRAM, enabling the vector database to launch instantly. This supports seamless switching between user-specific or application-specific databases on the same server for efficient RAG service delivery.
- Optimized for cloud systems by storing indexes in disaggregated storage for sharing across multiple servers. This approach dynamically adjusts vector database search performance for specific users or applications and facilitates the rapid migration of search instances between physical servers.

"The KIOXIA AiSAQ™ solution paves the way for almost infinite scaling of RAG applications in Generative AI Systems based of flash-based SSDs at the core," said Axel Stoermann, Chief Technology Officer & VP at KIOXIA Europe GmbH. "Utilizing SSD-based ANNS, we are reducing the reliance on costly DRAM while matching the performance needs of leading in-memory solutions – enhancing the performance range of large-scale RAG applications significantly."

KIOXIA is demonstrating its commitment to advancing AI by contributing its innovative [KIOXIA AiSAQ technology to the community as open-source software.](#)

# # #

Notes:

1: KIOXIA AiSAQ: All-in-Storage ANNS with Product Quantization, a novel method of index data placement, is a trademark of KIOXIA.

All other company names, product names and service names may be trademarks of third-party companies.

KIOXIA

About KIOXIA

KIOXIA is a world leader in memory solutions, dedicated to the development, production and sale of flash memory and solid-state drives (SSDs). In April 2017, its predecessor Toshiba Memory was spun off from Toshiba Corporation, the company that invented NAND flash memory in 1987. KIOXIA is committed to uplifting the world with "memory" by offering products, services and systems that create choice for customers and memory-based value for society. KIOXIA's innovative 3D flash memory technology, BiCS FLASH™, is shaping the future of storage in high-density applications, including advanced smartphones, PCs, automotive systems, data centers and generative AI systems.

Visit our [KIOXIA website](#)

Contact details for publication:
KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Germany
Tel: +49 (0)211 368 77-0
E-mail: KIE-support@kioxia.com

Contact details for editorial enquiries:
Lena Hoffmann, KIOXIA Europe GmbH
Tel: +49 (0) 211 36877 382
E-mail: lena1.hoffmann@kioxia.com

Issued by:
Birgit Schöniger, Publitek
Tel: +49 (0)172 617 8431
E-mail: birgit.schoeniger@publitek.com
Web: www.publitek.com