



Communiqué de presse

La technologie AiSAQ™ de KIOXIA est conçue pour réduire les besoins en DRAM dans les systèmes d'IA générative commercialisés en tant que logiciels open source

La technologie logicielle améliore la mise à l'échelle des bases de données vectorielles et la précision des flux de travail RAG à l'aide des SSD.

Allemagne, Düsseldorf, le 29 janvier 2025 – [KIOXIA](#) a annoncé aujourd'hui la sortie de sa nouvelle technologie AiSAQ (All-in-Storage ANNS with Product Quantization) [en open source](#). Avec un nouvel algorithme de recherche du plus proche voisin approximatif (ANNS) optimisé pour les disques SSD, le logiciel KIOXIA AiSAQ™⁽¹⁾ offre des performances évolutives pour la génération augmentée par récupération (RAG) sans placer les données d'index dans la DRAM – et effectuer à la place une recherche directement sur les disques SSD.

Les systèmes d'IA générative exigent d'importantes ressources de calcul, de mémoire et de stockage. Bien qu'ils aient le potentiel de générer des percées transformatrices dans divers secteurs, leur déploiement s'accompagne souvent de coûts élevés. La RAG est un élément critique de l'IA qui permet d'affiner les grands modèles de langage (LLM) à l'aide de données spécifiques à l'entreprise ou à l'application.

Un composant central de la RAG est une base de données vectorielle qui accumule et convertit des données spécifiques en vecteurs de caractéristiques dans la base de données. La RAG utilise également un ANNS qui identifie les vecteurs susceptibles d'améliorer le modèle en fonction de la similitude entre les vecteurs accumulés et les vecteurs cibles.



Pour que la RAG soit efficace, elle doit récupérer rapidement les informations les plus pertinentes pour une requête. Traditionnellement, les ANNS sont déployés dans la DRAM pour atteindre les performances à haut débit requises pour ces recherches.

La technologie AiSAQ™ de KIOXIA fournit une solution ANNS évolutive et efficace pour des ensembles de données à l'échelle d'un milliard avec une utilisation négligeable de la mémoire et des capacités de commutation d'index rapides.

Principaux avantages de la technologie KIOXIA AiSAQ™ :

- Permet aux bases de données à grande échelle de fonctionner sans dépendre de ressources DRAM limitées, ce qui améliore les performances des systèmes RAG.
- Il n'est plus nécessaire de charger les données d'index dans la DRAM, ce qui permet à la base de données vectorielle de se lancer instantanément. Cela permet de basculer de manière transparente entre des bases de données spécifiques à l'utilisateur ou à l'application sur le même serveur pour fournir efficacement un service RAG.
- La technologie est optimisée pour les systèmes cloud en stockant des index dans un stockage désagrégé pour le partage sur plusieurs serveurs. Cette approche ajuste de façon dynamique les performances de recherche de bases de données vectorielles pour des utilisateurs ou des applications spécifiques et facilite la migration rapide des instances de recherche entre les serveurs physiques.

« La solution AiSAQ™ de KIOXIA ouvre la voie à une évolutivité presque infinie des applications RAG dans les systèmes d'IA générative fondés sur des SSD basés sur la mémoire flash », a déclaré Axel Stoermann, directeur de la technologie et vice-président de KIOXIA Europe GmbH. « En utilisant l'ANNS basé sur SSD, nous réduisons la dépendance en matière de DRAM coûteuse, tout en répondant aux besoins de performance des principales solutions en mémoire, améliorant ainsi considérablement la gamme de performances des applications RAG à grande échelle. »

KIOXIA démontre son engagement à faire progresser l'IA en apportant sa [technologie innovante KIOXIA AiSAQ à la communauté sous forme de logiciel open source.](#)



#

Remarques :

1 : KIOXIA AiSAQ : l'ANNS tout-en-un avec quantification du produit, nouvelle méthode de placement des données d'index, est une marque déposée de KIOXIA.

Tous les autres noms de sociétés, noms de produits et noms de services peuvent être des marques commerciales de sociétés tierces.

À propos de KIOXIA

KIOXIA est un leader mondial des solutions de mémoire, dédié au développement, à la production et à la vente de mémoire flash et de disques d'état solide (SSD). En avril 2017, son prédécesseur Toshiba Memory a été séparé de Toshiba Corporation, la société qui a inventé la mémoire flash NAND en 1987. KIOXIA s'engage à améliorer le monde grâce à la "mémoire" en offrant des produits, des services et des systèmes qui créent un choix pour les clients et une valeur basée sur la mémoire pour la société. La technologie innovante de mémoire flash 3D de KIOXIA, BiCS FLASH™, façonne l'avenir du stockage dans les applications à haute densité, notamment les smartphones avancés, les PC, les systèmes automobiles, les centres de données et les systèmes d'IA générative.

Visitez le [site web de KIOXIA](#)

Coordonnées pour la publication :

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Allemagne

Tél. : +49 (0)211 368 77-0

E-mail : KIE-support@kioxia.com

Coordonnées pour les demandes d'information éditoriale :

Lena Hoffmann, KIOXIA Europe GmbH

Tél. : +49 (0) 211 36877 382

E-mail : lena1.hoffmann@kioxia.com

KIOXIA

Publié par :

Birgit Schöniger, Publitek

Tél. : +49 (0)172 617 8431

E-mail : birgit.schoeniger@publitek.com

Site web : www.publitek.com