



Comunicato stampa

La tecnologia KIOXIA AiSAQ™ è stata progettata per ridurre i requisiti di DRAM nei sistemi di intelligenza artificiale generativa rilasciata come software open source

La tecnologia software migliora la scalabilità e l'accuratezza dei database vettoriali nei flussi di lavoro RAG utilizzando gli SSD.

Germania, Düsseldorf, 29 gennaio 2025 – [KIOXIA](#) ha annunciato oggi il rilascio [open-source](#) della sua nuova tecnologia AiSAQ (All-in-Storage ANNS with Product Quantization). Il software KIOXIA AiSAQ™⁽¹⁾, un nuovo algoritmo di ricerca ANNS (Approximate Nearest Neighbor) ottimizzato per gli SSD, offre prestazioni scalabili per la generazione aumentata dal recupero (RAG) senza necessità di inserire i dati di indice nella DRAM e cercando direttamente sugli SSD.

I sistemi di intelligenza artificiale generativa richiedono risorse di elaborazione, memoria e archiviazione significative. Sebbene abbiano il potenziale per promuovere innovazioni trasformative in vari settori, la loro implementazione spesso comporta costi elevati. Il RAG è una fase critica dell'intelligenza artificiale che perfeziona i modelli linguistici di grandi dimensioni (LLM) con dati specifici per l'azienda o l'applicazione.

Un componente centrale del RAG è un database vettoriale che accumula e converte dati specifici in vettori di caratteristiche nel database. Il RAG utilizza anche un algoritmo ANNS, che identifica i vettori che migliorano il modello in base alla somiglianza tra i vettori accumulati e quelli target. Affinché il RAG sia efficace, deve recuperare rapidamente le informazioni più rilevanti per una query. Tradizionalmente, gli algoritmi ANNS vengono implementati nella DRAM per ottenere le prestazioni ad alta velocità richieste per queste ricerche.



La tecnologia KIOXIA AiSAQ™ fornisce una soluzione ANNS scalabile ed efficiente per set di dati su scala miliardaria con un utilizzo trascurabile della memoria e capacità di commutazione rapida dell'indice.

Principali vantaggi della tecnologia KIOXIA AiSAQ™:

- Consente ai database su larga scala di funzionare senza fare affidamento su risorse DRAM limitate, migliorando le prestazioni dei sistemi RAG.
- Elimina la necessità di caricare i dati dell'indice nella DRAM, consentendo l'avvio immediato del database vettoriale. Ciò supporta il passaggio senza soluzione di continuità tra database specifici dell'utente o dell'applicazione sullo stesso server per un'efficiente erogazione del servizio RAG.
- Ottimizzato per i sistemi cloud grazie all'archiviazione degli indici in un archivio disaggregato per la condivisione su più server. Questo approccio regola dinamicamente le prestazioni di ricerca dei database vettoriali per utenti o applicazioni specifici e facilita la migrazione rapida delle istanze di ricerca tra server fisici.

"La soluzione KIOXIA AiSAQ™ apre la strada a una scalabilità quasi infinita delle applicazioni RAG nei sistemi di intelligenza artificiale generativa basati su SSD basati su flash al centro", ha dichiarato Axel Stoermann, Chief Technology Officer & VP di KIOXIA Europe GmbH. "Utilizzando un ANNS basato su SSD, stiamo riducendo la dipendenza dalla costosa DRAM, soddisfacendo al contempo le esigenze di prestazioni delle principali soluzioni in-memory, migliorando significativamente la gamma di prestazioni delle applicazioni RAG su larga scala".

KIOXIA sta dimostrando il suo impegno per far progredire [l'intelligenza artificiale, contribuendo alla comunità con la sua innovativa tecnologia KIOXIA AiSAQ come software open source.](#)

###



Note:

1: KIOXIA AiSAQ: AiSAQ, un nuovo metodo di posizionamento dei dati indicizzati, è un marchio di fabbrica di KIOXIA.

Tutti gli altri nomi di società, nomi di prodotti e nomi di servizi possono essere marchi di società terze.

Informazioni su KIOXIA

KIOXIA è un leader mondiale nelle soluzioni di memoria, dedicato allo sviluppo, alla produzione e alla vendita di memorie flash e unità a stato solido (SSD). Nell'aprile 2017, il suo predecessore Toshiba Memory è stato scorporato da Toshiba Corporation, l'azienda che ha inventato la memoria flash NAND nel 1987. KIOXIA si impegna a migliorare il mondo con la "memoria", offrendo prodotti, servizi e sistemi che creano scelta per i clienti e valore basato sulla memoria per la società. L'innovativa tecnologia di memoria flash 3D di KIOXIA, BiCS FLASH™, sta plasmando il futuro dell'archiviazione nelle applicazioni ad alta densità, tra cui smartphone avanzati, PC, sistemi automobilistici, data center e sistemi di intelligenza artificiale generativa.

Visita il [sito web KIOXIA](#)

Contatti per la pubblicazione:

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Germania

Tel: +49 (0)211 368 77-0

E-mail: KIE-support@kioxia.com

Dettagli dei contatti per richieste editoriali:

Lena Hoffmann, KIOXIA Europe GmbH

Tel: +49 (0) 211 36877 382

E-mail: lena1.hoffmann@kioxia.com

Pubblicato da:

Birgit Schöniger, Publitek

Tel: +49 (0)172 617 8431

E-mail: birgit.schoeniger@publitek.com

Web: www.publitek.com