



Pressmeddelande

Programvaran KIOXIA AiSAQ förbättrar AI RAG med ny version av vektorsökning

Ny programvara med öppen källkod möjliggör flexibel balansering av kapacitet och prestanda baserat på användares behov och miljö

Tyskland, Düsseldorf, 3 juli 2025 – Inom ramen för det pågående arbetet att förbättra användbarheten av AI-vektordatabassökningar i RAG-system (Retrieval-Augmented Generation) genom optimering av användningen av solid-state-enheter (SSD), tillkännagav KIOXIA idag en uppdatering av sin programvara KIOXIA AiSAQ™ (All-in-Storage ANNS with Product Quantization). Denna nya version med öppen källkod introducerar flexibla kontroller som gör det möjligt för systemarkitekter att definiera balansen mellan sökprestanda och antalet vektorer, som är motsatta faktorer i den fasta kapaciteten för SSD-lagring i systemet. Det skapar fördelar som gör att arkitekter av RAG-system kan finjustera den optimala balansen mellan specifika arbetsbelastningar och deras krav, utan några hårdvaruändringar.

Programvaran KIOXIA AiSAQ, som [lanserades i januari 2025](#), använder en ny ANNS-algoritm (Approximate Nearest Neighbor Search) som är optimerad för SSD-enheter och eliminerar behovet av att lagra indexdata i DRAM. Genom att möjliggöra vektorsökningar direkt på SSD-enheter och minska kraven på värdminne gör KIOXIA AiSAQ-tekniken det möjligt för vektordatabaser att skalas, i stort sett utan de begränsningar som orsakas av begränsad DRAM-kapacitet.

När den installerade kapaciteten för SSD:n i systemet är fast kräver ökad sökprestanda (frågor per sekund) att mer SSD-kapacitet förbrukas per vektor. Detta resulterar i ett lägre antal vektorer. Omvänt, för att maximera antalet vektorer, måste SSD:ns kapacitetsförbrukning per vektor

minskas, vilket resulterar i lägre prestanda. Den optimala balansen mellan dessa två motsatta villkor varierar beroende på den specifika arbetsbelastningen. För att hitta rätt balans introducerar KIOXIA AiSAQ-programvaran flexibla konfigurationsalternativ. Den senaste uppdateringen gör det möjligt för administratörer att välja den optimala balansen för en mängd olika arbetsbelastningar i RAG-systemet. Genom den här uppdateringen blir KIOXIA AiSAQ-tekniken en lämplig SSD-baserad ANNS för RAG-applikationer samt även för andra "vektorhungriga" applikationer, som semantiska offlinesökningar.

Med växande efterfrågan på skalbara AI-tjänster erbjuder SSD-enheter ett praktiskt alternativ till DRAM för att hantera den höga genomströmning och låga latens som RAG-system kräver. Genom KIOXIA AiSAQ-programvaran kan man möta dessa krav på ett effektivt sätt, vilket ger möjligheter till storskalig generativ AI som inte hämmas av begränsade minnesresurser.

Eftersom programvaran KIOXIA AiSAQ släpps med öppen källkod förstärker KIOXIA sitt engagemang inom AI-communityn genom att främja SSD-centrerade arkitekturer för skalbar AI.

"Vi försöker alltid hjälpa utvecklare och systemarkitekter att finjustera prestanda och kapacitet på nya innovativa sätt", säger Axel Störmann, Vice President och Chief Technology Officer för minnes- och SSD-produkter på KIOXIA Europe GmbH. "Nu, med introduktionen av den senaste versionen av KIOXIA AiSAQ-programvaran, kan slutanvändare lita på kraften hos SSD-enheter för att bygga skalbara RAG-system flexibelt och effektivt. Genom att använda öppen källkod för vår teknik bekräftar vi vårt orubbliga engagemang för AI-communityn och ger dem lösningar som är både kraftfulla och tillgängliga."

Följ länken för att ladda ner programvaran KIOXIA AiSAQ med öppen källkod.

<https://github.com/kioxia-jp/aisaq-diskann>

###

Fotnoter:

*KIOXIA AiSAQ: All-in-Storage ANNS with Product Quantization, en ny metod för placering av indexdata, är ett varumärke som tillhör KIOXIA.

*Alla andra namn på företag, produkter och tjänster kan vara varumärken som tillhör tredjepartsföretag.

Om KIOXIA

KIOXIA är världsledande inom minneslösningar och utvecklar, producerar och säljer flashminnen och SSD-enheter (Solid State Drive). I april 2017 avknoppades dess föregångare Toshiba Memory från Toshiba Corporation, företaget som uppfann NAND-flashminnet 1987. KIOXIA har som mål att förbättra världen med "minnen" genom att erbjuda produkter, tjänster och system som skapar valmöjligheter för kunderna och minnesbaserat värde för samhället. KIOXIAs innovativa 3D-flashminnesteknik, BiCS FLASH™ formar framtidens lagring med hög densitet för användning i bland annat avancerade smarttelefoner, datorer, fordonssystem, datacenter och generativa AI-system.

Besök [KIOXIAs webbplats](#)

Kontaktuppgifter för publicering:

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Tyskland

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

E-post: KIE-support@kioxia.com

Kontaktuppgifter för redaktionella förfrågningar:

Lena Hoffmann, KIOXIA Europe GmbH

Tel: +49 (0) 211 36877 382

E-post: lena1.hoffmann@kioxia.com

Publicerat av:

Birgit Schöniger, Publitek

Tel: +49 (0)172 617 8431

E-post: birgit.schoeniger@publitek.com

Webb: www.publitek.com