



Komunikat prasowy

Technologia KIOXIA AiSAQ™ zaprojektowana w celu zmniejszenia wymagań dotyczących pamięci DRAM w generatywnych systemach sztucznej inteligencji wydana jako oprogramowanie typu open source

Technologia oprogramowania poprawia skalowanie wektorowych baz danych i dokładność w przepływach pracy RAG dzięki wykorzystaniu dysków SSD.

Niemcy, Düsseldorf, 29 stycznia 2025 r. – Firma [KIOXIA](#) ogłosiła dziś wydanie wersji [open-source](#) nowej technologii AiSAQ (All-in-Storage ANNS z kwantyzacją produktu). Nowatorski algorytm ANNS (poszukiwania przybliżonego najbliższego sąsiada) zoptymalizowany dla dysków SSD w oprogramowaniu KIOXIA AiSAQ™⁽¹⁾ zapewnia skalowalne działanie generowania wspomaganego odzyskiwaniem (RAG) bez umieszczania danych indeksowania w pamięci DRAM, zamiast tego wyszukując bezpośrednio na dyskach SSD.

Systemy generatywnej sztucznej inteligencji wymagają znacznych zasobów obliczeniowych i pamięci, w tym masowej. Chociaż potencjalnie mogą przyczynić się do transformacyjnych przełomów w różnych branżach, ich wprowadzanie często wiąże się z wysokimi kosztami. Technika RAG to krytyczny etap dla sztucznej inteligencji, który udoskonala duże modele językowe (LLM) za pomocą danych specyficznych dla danej firmy lub zastosowania.

Centralnym komponentem techniki RAG jest wektorowa baza danych, która gromadzi i przekształca określone dane w wektory cech w bazie danych. Technika RAG wykorzystuje również algorytm ANNS, który identyfikuje wektory, które ulepszają model na podstawie podobieństwa między wektorami zebranymi i docelowymi.



Aby technika RAG była skuteczna, musi szybko pobierać informacje najbardziej istotne dla zapytania. Tradycyjnie, algorytmy ANNS są umieszczane w pamięci DRAM w celu osiągnięcia wysokiej wydajności wymaganej do tych wyszukiwań.

Technologia KIOXIA AiSAQ™ zapewnia skalowalne i wydajne rozwiązanie ANNS dla zestawów danych o skali miliarda przy znikomym zużyciu pamięci i możliwościach szybkiego przełączania indeksu.

Najważniejsze zalety technologii KIOXIA AiSAQ™:

- Umożliwia obsługę dużych baz danych bez polegania na ograniczonych zasobach pamięci DRAM, co zwiększa wydajność systemów RAG.
- Eliminuje konieczność wczytywania danych indeksu do pamięci DRAM, umożliwiając natychmiastowe uruchomienie wektorowej bazy danych. Dzięki temu można bezproblemowo przełączać się między bazami danych specyficznymi dla użytkownika lub zastosowania na tym samym serwerze w celu wydajnego świadczenia usług RAG.
- Rozwiązanie zoptymalizowano pod kątem systemów w chmurze dzięki przechowywaniu indeksów w zdezagregowanej pamięci masowej, co umożliwia udostępnianie na wielu serwerach. Takie dynamiczne podejście dostosowuje wydajność wyszukiwania w wektorowej bazie danych dla określonych użytkowników lub zastosowań i ułatwia szybką migrację instancji wyszukiwania między serwerami fizycznymi.

„Rozwiązanie KIOXIA AiSAQ™ toruje drogę do niemal nieograniczonego skalowania zastosowań RAG w generatywnych systemach sztucznej inteligencji wykorzystujących dyski SSD oparte na pamięci flash” – mówi Axel Stoermann, dyrektor ds. technologii i wiceprezes w KIOXIA Europe GmbH. „Korzystając z systemów ANNS wykorzystujących dyski SSD, zmniejszamy zależność od kosztownych pamięci DRAM, jednocześnie dopasowując się do potrzeb w zakresie wydajności wiodących rozwiązań zintegrowanych typu in-memory – co znacznie zwiększając zakres wydajności zastosowań RAG na dużą skalę”.

KIOXIA demonstruje swoje zaangażowanie w rozwój sztucznej inteligencji, udostępniając społeczności swoją innowacyjną [technologię KIOXIA AiSAQ jako oprogramowanie typu open source.](#)



#

Uwagi:

1: KIOXIA AiSAQ: All-in-Storage ANNS with Product Quantization — czyli nowatorska metoda umieszczania danych indeksowych — jest znakiem towarowym firmy KIOXIA.

Wszelkie pozostałe nazwy firm, produktów oraz usług mogą stanowić znaki towarowe podmiotów trzecich.

Informacje o KIOXIA Europe GmbH

KIOXIA jest światowym liderem w dziedzinie rozwiązań pamięciowych, zajmującym się rozwojem, produkcją i sprzedażą pamięci flash i dysków półprzewodnikowych (SSD). W kwietniu 2017 roku jej poprzednik, Toshiba Memory, został wydzielony z Toshiba Corporation, firmy, która wynalazła pamięć flash NAND w 1987 roku. KIOXIA jest zaangażowana w ulepszanie świata za pomocą "pamięci" poprzez oferowanie produktów, usług i systemów, które zapewniają klientom wybór, a społeczeństwu wartość opartą na pamięci. Innowacyjna technologia pamięci flash 3D firmy KIOXIA, BiCS FLASH™, kształtuje przyszłość pamięci masowej w zastosowaniach o dużej gęstości, w tym w zaawansowanych smartfonach, komputerach PC, systemach motoryzacyjnych, centrach danych i generatywnych systemach sztucznej inteligencji.

Odwiedź naszą [witrynę KIOXIA](#)

Dane kontaktowe do publikacji:

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Niemcy

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

E-mail: KIE-support@kioxia.com

Dane kontaktowe w przypadku zapytań redakcyjnych:

Lena Hoffmann, KIOXIA Europe GmbH

Tel: +49 (0) 211 36877 382

E-mail: lena1.hoffmann@kioxia.com

KIOXIA

Wydawca:

Birgit Schöniger, Publitek

Tel.: +49 (0)172 617 8431

E-mail: birgit.schoeniger@publitek.com

Strona internetowa: www.publitek.com