



## Пресс-релиз

**Технология KIOXIA AiSAQ™, разработанная для снижения требований к DRAM для генеративных систем ИИ, представлена в виде программного обеспечения с открытым исходным кодом**

*Благодаря использованию твердотельных накопителей технология программного обеспечения позволяет повысить масштабируемость и точность векторных баз данных в рабочих процессах RAG.*

Германия, Дюссельдорф, 29 января 2025 г. Компания [KIOXIA](#) сегодня объявила о выпуске [открытого исходного кода](#) технологии All-in-Storage ANNS with Product Quantization (AiSAQ). Новый алгоритм поиска «приблизительного ближайшего соседа» (ANNS), оптимизированный для твердотельных накопителей, программное обеспечение KIOXIA AiSAQ™<sup>(1)</sup> обеспечивает масштабируемую производительность для генерации с дополненной выборкой (RAG), выполняя поиск непосредственно на твердотельных накопителях без размещения индексных данных в DRAM.

Системы генеративного искусственного интеллекта требовательны к объемам оперативной памяти, вычислительным ресурсам и ресурсам хранения данных. Их внедрение зачастую сопряжено с высокими затратами, несмотря на потенциал, позволяющий совершать революционные прорывы в различных отраслях. RAG — это критический этап ИИ, на котором происходит уточнение больших языковых моделей (LLM) на основе данных, специфичных для компании или приложения.



Центральным компонентом RAG является векторная база данных, которая накапливает и преобразует конкретные данные в векторы признаков в базе данных. В RAG также используется алгоритм ANNS, который на основании сходства между накопленными и целевыми векторами определяет векторы, улучшающие модель. Чтобы RAG был эффективным, он должен быстро находить информацию, наиболее релевантную запросу. Для достижения высокой скорости, необходимой для такого поиска, алгоритмы ANNS традиционно выполняются в DRAM.

Технология KIOXIA AiSAQ™ обеспечивает более экономичное и масштабируемое решение ANNS для миллиардно-масштабируемых массивов данных с незначительным использованием памяти и возможностью быстрого переключения индексов.

#### **Основные преимущества технологии KIOXIA AiSAQ™:**

- обеспечивает работу крупномасштабных баз данных без зависимости от ограниченных ресурсов DRAM, повышая производительность систем RAG;
- позволяет мгновенно запускать векторную базу данных за счет устранения необходимости загрузки индексных данных в DRAM; обеспечивает беспрепятственное переключение между базами данных для конкретных пользователей или приложений на одном сервере для эффективного предоставления сервисов RAG;
- оптимизирована для облачных систем за счет хранения индексов в деагрегированном хранилище для совместного использования между множеством серверов. Такой подход позволяет динамически регулировать производительность векторного поиска в базе данных для конкретных пользователей или приложений и облегчает быструю миграцию поисковых экземпляров между физическими серверами.

«Решение KIOXIA AiSAQ™ открывает путь к практически бесконечному масштабированию приложений RAG в генеративных системах искусственного интеллекта, основанных на твердотельных накопителях на базе флэш-памяти, — сообщил Аксель Штёрманн (Axel Stoermann), директор по технологиям и вице-президент компании KIOXIA Europe GmbH. —



Используя ANNS на основе твердотельных накопителей (SSD), мы снижаем зависимость от дорогостоящей DRAM, обеспечивая при этом потребности в производительности ведущих решений в области оперативной памяти, значительно расширяя диапазон производительности крупномасштабных приложений RAG».

KIOXIA демонстрирует свою приверженность продвижению ИИ, предлагая сообществу свою инновационную технологию [KIOXIA AiSAQ в виде программного обеспечения с ОТКРЫТЫМ ИСХОДНЫМ КОДОМ.](#)

# # #

**Примечания:**

1. KIOXIA AiSAQ: All-in-Storage ANNS with Product Quantization (AiSAQ) — новый метод размещения индексных данных, является торговой маркой компании KIOXIA.

Все другие названия компаний, названия продуктов и услуг могут быть товарными знаками сторонних компаний.

**О компании KIOXIA Europe GmbH**

KIOXIA Europe GmbH (ранее Toshiba Memory Europe GmbH) — расположенное в Европе подразделение KIOXIA Corporation, ведущего мирового поставщика флэш-памяти и твердотельных накопителей (SSD). Начиная с изобретения флэш-памяти NAND и заканчивая хорошо известной сегодня 3D флэш памятью BiCS FLASH™, KIOXIA продолжает внедрять передовые решения и услуги в области памяти, которые улучшают жизнь людей и расширяют возможности общества. Инновационная технология флэш-памяти BiCS FLASH™ 3D определяет будущее систем хранения данных в приложениях с высокой плотностью, включая современные смартфоны, ПК, твердотельные накопители (SSD), автомобильные и дата-центры, а также системы генеративного ИИ.

Посетите [веб-сайт KIOXIA](#)

**Контактные данные для публикации:**

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Germany (г. Дюссельдорф, Германия)

Тел.: +49 (0)211 368 77-0

Электронная почта: [KIE-support@kioxia.com](mailto:KIE-support@kioxia.com)