



Basın Bülteni

KIOXIA AiSAQ™ Teknolojisi, Üretken Yapay Zeka Sistemlerinde DRAM Gereksinimlerini Azaltmak için Tasarlanan Açık Kaynak Kodlu Yazılım Olarak Yayınlandı

Yazılım teknolojisi, SSD'leri kullanarak RAG iş akışlarında vektör veritabanı ölçeklendirmesini ve doğruluğunu geliştiriyor.

Almanya, Düsseldorf, 29 Ocak 2025 – KIOXIA, bugün yeni All-in-Storage ANNS with Product Quantization (AiSAQ) teknolojisinin [açık kaynaklı](#) sürümünü duyurdu. SSD'ler için optimize edilmiş yeni bir “yaklaşık en yakın komşu” arama (ANNS) algoritması olan KIOXIA AiSAQ™⁽¹⁾ yazılımı, dizin verilerini DRAM'e yerleştirmeden ve bunun yerine doğrudan SSD'lerde arama yaparak erişim artırımı nesil (RAG) için ölçeklenebilir performans sunuyor.

Üretken yapay zeka sistemleri önemli miktarda hesaplama, bellek ve depolama kaynağı gerektiriyor. Çeşitli sektörlerde dönüştürücü atılımlar gerçekleştirme potansiyeline sahip olsalar da, dağıtımları genellikle yüksek maliyetlerle birlikte geliyor. RAG, büyük dil modellerini (LLM'ler) şirkete veya uygulamaya özel verilerle iyileştiren kritik bir yapay zeka aşaması olarak öne çıkıyor.

RAG'nin merkezi bir bileşeni, belirli verileri biriktiren ve veritabanında özellik vektörlerine dönüştüren bir vektör veritabanı. RAG ayrıca, birikmiş ve hedef vektörler arasındaki benzerliğe dayalı olarak modeli iyileştiren vektörleri tanımlayan bir ANNS algoritması kullanıyor. RAG'nin etkili olabilmesi için, bir sorguyla en alakalı bilgileri hızlı bir şekilde alması gerekiyor. Geleneksel olarak ANNS algoritmaları, bu aramalar için gereken yüksek hızlı performansı elde etmek için DRAM'de kullanılıyor.



KIOXIA AiSAQ™ teknolojisi, ihmal edilebilir bellek kullanımı ve hızlı dizin deęiřtirme özellikleriyle milyar ölçekli veri kümeleri için ölçeklenebilir ve verimli bir ANNS çözümü sunuyor.

KIOXIA AiSAQ™ Teknolojisinin Temel Faydaları:

- Büyük ölçekli veritabanlarının sınırlı DRAM kaynaklarına ihtiyaç duymadan çalışmasını sağlayarak RAG sistemlerinin performansını artırıyor.
- Dizin verilerini DRAM'e yükleme ihtiyacını ortadan kaldırarak vektör veritabanının anında başlatılmasını sağlıyor. Bu, verimli RAG hizmeti sunumu için aynı sunucu üzerinde kullanıcıya özel veya uygulamaya özel veritabanları arasında sorunsuz geçiři destekliyor.
- Dizinleri birden fazla sunucu arasında paylaşmak için ayrıştırılmış depolama alanında depolayarak bulut sistemleri için optimize edildi. Bu yaklaşım, belirli kullanıcılar veya uygulamalar için vektör veritabanı arama performansını dinamik olarak ayarlıyor ve fiziksel sunucular arasında arama örneklerinin hızlı geçiřini kolaylaştırıyor.

KIOXIA Europe GmbH Baş Teknoloji Sorumlusu ve Başkan Yardımcısı Axel Stoermann, “KIOXIA AiSAQ™ çözümü, temelde flař tabanlı SSD'lere dayanan Generative AI Sistemlerinde RAG uygulamalarının neredeyse sonsuz ölçeklendirilmesinin önünü açıyor” dedi ve ekledi: “SSD tabanlı ANNS kullanarak, önde gelen bellek içi çözümlerin performans ihtiyaçlarını karşılarken maliyetli DRAM'e olan bağımlılıęı azaltıyoruz ve büyük ölçekli RAG uygulamalarının performans aralıęını önemli ölçüde artırıyoruz.”

KIOXIA, yenilikçi [KIOXIA AiSAQ teknolojisini açık kaynaklı yazılım olarak topluma sunarak yapay zekayı geliştirme konusundaki kararlılıęını ortaya koyuyor.](#)

###

KIOXIA

Notlar:

1: KIOXIA AiSAQ: Yeni bir dizin verisi yerleřtirme yöntemi olan Ürün Nicelemeli All-in-Storage ANNS, KIOXIA'nın ticari markasıdır.

Diđer tüm řirket adları, ürün adları ve hizmet adları üçüncü taraf řirketlerin ticari markaları olabilir.

KIOXIA Hakkında

KIOXIA, flash bellek ve katı hal sürücülerinin (SSD'ler) geliştirilmesi, üretimi ve satışına adanmış bellek çözümlerinde bir dünya lideridir. Nisan 2017'de, selefi Toshiba Memory, 1987 yılında NAND flash belleği icat eden Toshiba Corporation'dan ayrılmıştır. KIOXIA, müşteriler için seçenek ve toplum için bellek tabanlı değer yaratan ürünler, hizmetler ve sistemler sunarak dünyayı "bellek" ile yükseltmeye kararlıdır. KIOXIA'nın yenilikçi 3D flash bellek teknolojisi BiCS FLASH™, gelişmiş akıllı telefonlar, PC'ler, otomotiv sistemleri, veri merkezleri ve üretken yapay zeka sistemleri dahil olmak üzere yüksek yoğunluklu uygulamalarda depolamanın geleceğini şekillendiriyor.

[KIOXIA web sitemizi](#) ziyaret edin

Yayın için iletişim bilgileri:

KIOXIA Europe GmbH, Hansaallee 181, 40549 Düsseldorf, Almanya

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

E-posta: KIE-support@kioxia.com

Editöryel sorular için iletişim bilgileri:

Lena Hoffmann, KIOXIA Europe GmbH

Telefon: +49 (0) 211 36877 382

E-posta: lena1.hoffmann@kioxia.com

Hazırlayan:

Birgit Schöniger, Publitek

Tel: +49 (0)172 617 8431

E-posta: birgit.schoeniger@publitek.com

Web sitesi: www.publitek.com