

Dnotitia and HyperAccel Collaborate to Develop Optimized AI Inference System for RAG

Dnotitia and HyperAccel partner to integrate their AI semiconductor chips, delivering the world's first AI inference solution specifically optimized for RAG



Dnotitia x HyperAccel logo

SEOUL, SOUTH KOREA, March 20, 2025

/EINPresswire.com/ -- [Dnotitia](#), Inc.

(Dnotitia), a startup specializing in integrated artificial intelligence (AI) and semiconductor solutions, today announced a strategic partnership with [HyperAccel](#), a fabless semiconductor startup focusing on AI acceleration, to jointly develop an AI inference system optimized to Retrieval-Augmented Generation(RAG). This collaboration will integrate Dnotitia's Vector Data Processing Unit (VDPU) chip with HyperAccel's Large Language Model (LLM) accelerator chip, known as the LLM Processing Unit (LPU), into a unified system.

As data retrieval becomes increasingly crucial in AI services, with expanding data volumes and diverse data modalities, there is a rising demand for faster and more efficient data retrieval capabilities. Traditional systems rely heavily on software-based retrieval and handle LLM-based GenAI processes separately, leading to slower response times and higher power consumption. Dnotitia enables real-time retrieval and utilization of large-scale multimodal data by leveraging its VDPU, while HyperAccel maximizes AI model performance with its LPU chip. By combining these two technologies and optimizing at the system level, the companies aim to create the world's first RAG-specialized AI system capable of handling both retrieval and inference simultaneously.

"As LLM services become widespread, the demand for data retrieval is also rapidly increasing," said Moo-Kyoung (MK) Chung, CEO of Dnotitia. "Through this collaboration, we will introduce a new concept of AI system that not only optimizes the inference of AI models but also streamlines data retrieval. By applying long-term memory to AI, we can gain a deeper understanding of user data and provide more precise, customized services, and this will reduce hallucinations and serve as a key milestone toward more specialized and personalized AI services"

"Addressing computational bottlenecks while simultaneously improving performance and efficiency is the core challenge in AI semiconductor innovation," said Joo-Young Kim, CEO of

HyperAccel. "Our partnership seeks to introduce an optimized AI system tailored specifically for RAG and LLM applications, setting a critical milestone that will revolutionize how AI systems operate."

Dnotitia Communications

Dnotitia, Inc.

+82 2-588-1002

[email us here](#)

Visit us on social media:

[Facebook](#)

[LinkedIn](#)

[YouTube](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/795445280>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.