



Blaize and OrionVM to Deliver Hosted AI Inferencing Environment

September 27, 2024

Blaize GSP to power cloud solutions in video surveillance, facial recognition, medical imaging, and retail

San Francisco, Calif., and the AI INFRASummit, September 27, 2024 7:00 AM PT — [OrionVM®](#), a next-generation Cloud and AI company, today announced a further evolution of their strategic partnership with [Blaize®](#), a leader of purpose-built, AI-enabled edge computing solutions, to launch a hosted Artificial Intelligence (AI) inferencing environment powered by Blaize's Graph Streaming Processor (GSP®). Through this collaboration, Blaize's cutting-edge AI hardware and software offerings are now available in virtualized instances via the OrionVM cloud platform, unlocking new levels of scalability, accessibility, and performance for businesses, independent software vendors (ISVs), and AI developers. This new offering delivers an efficient, cost-effective alternative to traditional GPU-based systems for developing AI applications in video surveillance, facial recognition, medical imaging, retail sectors, and more.

Blaize GSP technology, designed specifically for AI inferencing, is a lower-power, lower-cost alternative to GPUs. Combining Blaize's high-efficiency GSPs with OrionVM's high-performance InfiniBand networking creates an ideal environment for low-latency, high-throughput AI and data processing workloads. The multi-tiered architecture of OrionVM allows for flexible resource allocation and efficient scaling, enabling companies to optimize their infrastructure costs while maintaining performance. Additionally, OrionVM's white-labeling capability allows service providers to customize and brand their offerings, extending tailored solutions to customers while leveraging the power and efficiency of Blaize GSPs hosted on OrionVM's infrastructure.

This integration provides a comprehensive, performance-driven cloud solution for enterprises handling complex AI, analytics, and data-intensive tasks. Customers can now access Blaize's hardware and software directly from the cloud, removing traditional barriers to adoption, such as hardware procurement and setup time, and significantly reducing time-to-market for AI-driven solutions.

Key Features of the Partnership:

- **Cloud-based AI Hardware:** Blaize's powerful edge computing devices, such as the Blaize Pathfinder and Xplorer platforms, will be available as cloud-hosted services, enabling users to leverage the full capabilities of AI processing without requiring physical hardware on-site.
- **Software-Driven AI Acceleration:** Blaize's cutting-edge AI software development platform, including the open-source Blaize AI Studio, will fully integrate into OrionVM, offering users a comprehensive toolset to develop, test, and deploy AI applications directly from the cloud.
- **Scalability and Flexibility:** OrionVM's highly efficient and flexible cloud infrastructure allows businesses to dynamically scale their AI operations according to demand, providing on-demand resources for large-scale AI projects and pay-as-you-go pricing models. OrionVM's robust cloud infrastructure, combined with GSP inferencing capabilities, ensures that applications can be deployed with low latency and high performance worldwide, providing ISVs the flexibility to grow and adapt as business needs evolve.
- **Privacy/Security.** Blaize and OrionVM offer a secure, private environment for AI workloads. Allowing clients to retain full control of their data, models, and output using Blaize GSP on an OrionVM private cloud deployment without requiring disclosure to expensive, limited public models. OrionVM's product suite is a one-stop shop for hosting and securing every component along an AI journey.

The new offering also empowers industry-specific solutions, including:

- **Video Surveillance:** Real-time AI analytics for monitoring, threat detection, and behavior analysis.
- **Facial Recognition:** AI-powered security and operational enhancements with automated identity verification.
- **Medical Imaging:** Enhanced diagnostic accuracy and advanced imaging for healthcare professionals, accelerating patient care and outcomes.
- **Smart Retail:** Intelligent AI systems that enhance customer experiences through facial recognition, product recommendations, and in-store analytics.

"Our partnership with Blaize introduces a breakthrough in AI application development for ISVs," said Daniel Pfeiffer, COO & VP Partnerships at OrionVM. "By combining our cloud infrastructure with Blaize's GSPs, we're enabling developers to deliver AI-driven solutions faster and more affordable than ever before. This collaboration will bring tremendous value to industries relying on real-time inferencing."

“We are thrilled to partner with OrionVM to bring Blaize’s AI solutions into the cloud,” said Dinakar Munagala, CEO of Blaize. “This collaboration represents a major milestone in our mission to democratize access to AI technology, enabling companies of all sizes to deploy and scale AI-powered applications more easily and cost-effectively than ever before.”

The joint offering is now available by contacting the OrionVM sales team. For more information, please visit <https://www.orionvm.com/blaizeai/>

About OrionVM

OrionVM is a leading Cloud and AI provider, offering high-performance cloud infrastructure to simplify and accelerate the deployment of modern applications. With a focus on scalability, performance, and cost-efficiency, OrionVM empowers developers and businesses to build, scale, and manage their AI and cloud workloads with ease. For more information, visit www.orionvm.com.

About Blaize

Blaize provides a full-stack programmable processor architecture suite and low-code/no-code software platform that enables AI processing solutions for high-performance computing at the network’s edge and in the data center. Blaize solutions deliver real-time insights and decision-making capabilities at low power consumption, high efficiency, minimal size, and low cost. Blaize has raised over \$330 million from strategic investors such as DENSO, Mercedes-Benz AG, Magna, and Samsung and financial investors such as Franklin Templeton, Temasek, GGV, Bess Ventures, BurTech LP LLC, Rizvi Traverse, and Ava Investors. Headquartered in El Dorado Hills (CA), Blaize has more than 200 employees worldwide with teams in San Jose (CA), Cary (NC), and subsidiaries in Hyderabad (India), Leeds and Kings Langley (UK), and Abu Dhabi (UAE).