

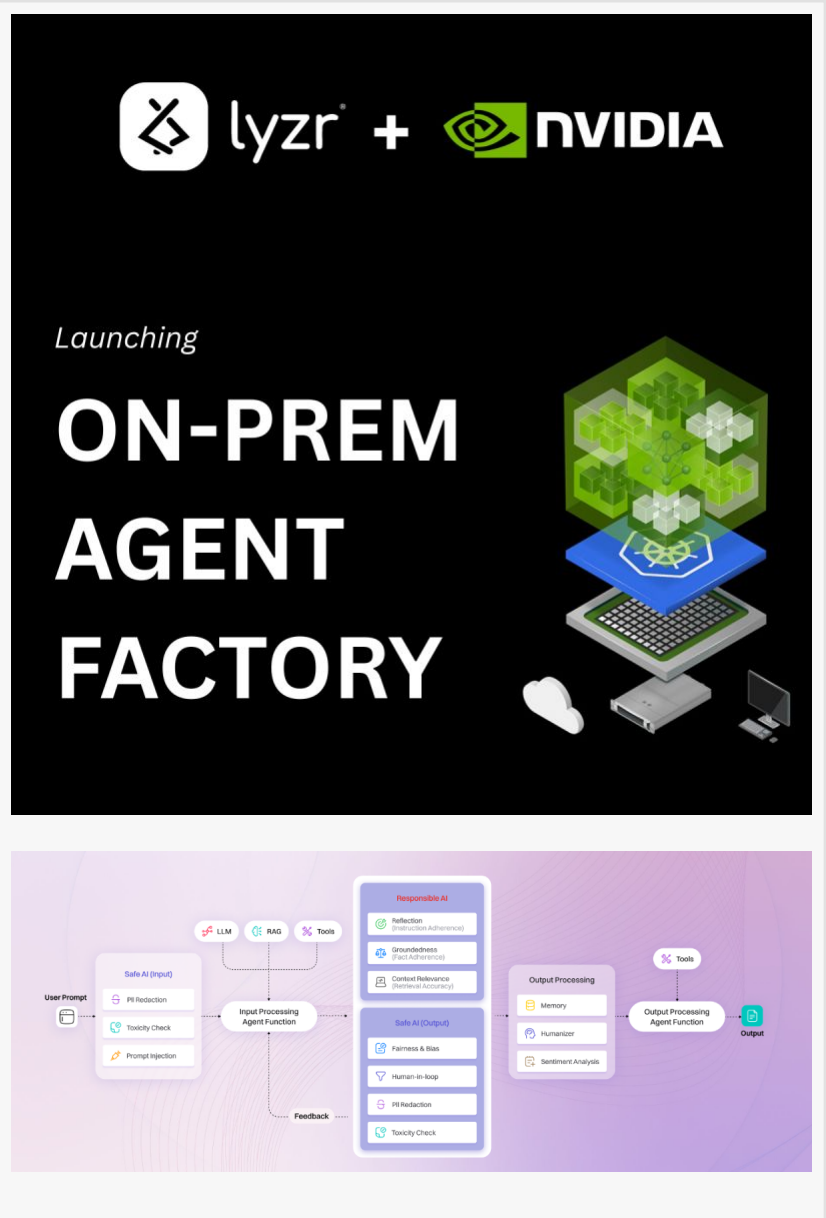
Lyzir AI Launches On-Prem Agent Factory Powered by NVIDIA NIMs for Financial Services

NEW JERSEY, NJ, UNITED STATES, April 24, 2025 /EINPresswire.com/ --

[Lyzir](#) AI has announced support for NVIDIA NIMs (NVIDIA Inference Microservices), allowing enterprises to build and manage AI agents entirely within their own infrastructure. This new integration is designed for organizations that already run on NVIDIA GPU clusters, especially those in financial services, insurance, and investment sectors, offering a secure, high-performance way to build AI systems on-premises.

NVIDIA NIMs are a suite of pre-configured AI models built for easy deployment on NVIDIA hardware. These models are optimized for fast inference and can be run entirely on local infrastructure. With [Lyzir Agent Studio](#) now supporting NIMs, companies can create and manage large-scale agent systems without relying on external servers or cloud-based platforms. This means data stays within the organization, addressing growing concerns around security, privacy, and regulatory compliance.

For banks, insurance providers, and hedge funds that have already invested heavily in NVIDIA GPU clusters, this partnership brings a practical way to make full use of existing hardware. Lyzir's



new enterprise offering enables these organizations to create what it calls an "On-Prem Agent Factory", a private environment for building and managing agents and multi-agent systems. These agents can automate internal operations, support research and analysis, and power customer-facing applications, all while running entirely within the organization's own infrastructure.

A Practical Path to Enterprise-Ready AI

NVIDIA NIMs are a set of containerized AI microservices that support fast and efficient inference using NVIDIA hardware. Each NIM is packaged as a Docker container containing a specific model—such as meta/llama3-8b-instruct—along with a runtime environment. These containers are designed to run on any NVIDIA GPU with adequate memory, and some model/GPU combinations are further optimized for performance. NIMs automatically retrieve models from NVIDIA NGC, using a local cache when available. Once the base container is downloaded, subsequent downloads are significantly faster due to a shared foundation across NIMs.

This architecture allows enterprises to deploy AI capabilities at scale while using their existing NVIDIA infrastructure. NIMs can be launched in seconds and are tuned for low-latency, high-throughput performance, making them well-suited for regulated industries with strict infrastructure and data requirements. By supporting NIMs, Lyzr ensures that enterprises can build agents with high-performance models, while maintaining compliance with internal data policies.

The demand for on-premise AI solutions is growing, particularly among companies that handle sensitive financial data. With this launch, Lyzr addresses a key gap: how to bring the flexibility of AI agents to highly regulated industries without compromising security.

This makes it possible for developers, analysts, and operations teams to build intelligent systems faster, using the infrastructure they already trust.

Lyzr's integration with NVIDIA NIMs is available now as part of its Enterprise Edition.

Organizations interested in exploring the platform can schedule a [demo](#) to see how the On-Prem Agent Factory works in practice and how it can support internal AI initiatives. For companies looking to move quickly with AI, without moving their data, this new partnership offers a secure, efficient, and scalable path forward.

About Lyzr

Founded in April 2023, Lyzr is dedicated to helping enterprises build and deploy reliable AI agents with a strong focus on privacy, scalability, and seamless integration. Backed by institutional investors, Lyzr serves over 400 clients, key clients include, AirAsia, Nelson Global, Accenture, Dairyland Power, and Marubeni. Its solutions portfolio spans AI and vertical AI agents tailored for banking and insurance, driving efficiency in automation and decision-making processes. Lyzr is headquartered in New York, with locations in San Francisco and Bangalore.

Anirudh Narayan

Lyzr AI

[email us here](#)

Visit us on social media:

[LinkedIn](#)

[YouTube](#)

[X](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/806269755>

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.