

July 16, 2026

Mizuho Financial Group, Inc.

Mizuho Begins Exploring Advanced AI Platforms for Financial Institutions Using NVIDIA Technologies

–Using DGX B200 and NemoClaw to strengthen computing and validate secure AI agent execution–

Mizuho Financial Group, Inc. (President & Group CEO: Masahiro Kihara; hereinafter, “Mizuho”) has begun exploring ways to advance of its AI infrastructure by leveraging NVIDIA's products, services, and technical expertise, with the aim of promoting more secure and sophisticated use of generative AI and AI agents by financial institutions.

As part of this study, Mizuho will examine key technical considerations from two perspectives: an on-premises GPU environment that supports R&D and inference validation for generative AI, and a framework for securely executing and managing AI agents. Through this initiative, Mizuho aims to expand the use of AI while ensuring the high standards of security and governance required at financial institutions.

By doing so, Mizuho seeks to appropriately protect highly confidential information while gradually expanding the scope of tasks supported by AI agents for employees, including information gathering, document preparation, analysis, and development support.

Background and Objectives

Generative AI and AI agents are expected to be applied across a wide range of operations at financial institutions, including information gathering, document preparation, inquiry handling, sales support, credit assessment and monitoring, and system development support.

At the same time, in order for financial institutions to fully utilize these technologies, they must meet stringent criteria: protection of highly confidential data and business information, compliance with laws, regulations, and internal rules, auditability of usage, and control over agent execution environments. It is essential to balance the expansion of AI use with the strict governance and control required in the financial sector.

In the future, AI agents are expected to operate continuously in response to triggers such as

emails, chats, notifications from business systems, and recurring business events, autonomously supporting tasks such as information gathering, document drafting, confirmation requests, and process execution.

To promote such use in financial institutions, it is essential to build an environment in which confidential information is not exposed externally, agent permissions and destinations can be controlled, and execution histories can be reviewed. With support from NVIDIA, Mizuho will examine the suitability and reliability of an AI infrastructure tailored to financial institutions from both the perspective of an on-premises GPU environment and a secure AI agent execution environment.

Overview of the Initiative

As part of this study, Mizuho will focus on the following two areas:

1. Technical considerations toward the introduction of NVIDIA DGX B200 and the future construction of a GPU cluster

As part of strengthening its on-premises GPU environment, Mizuho will move forward with the introduction of NVIDIA DGX B200^{*1}. Mizuho will use this environment for the training, evaluation, improvement, and inference validation of generative AI models, including the Mizuho LLM^{*2} currently under development, while also preparing an R&D environment for proof-of-concept use case testing.

In addition, with a view to future expansion of AI use and increasing inference demand, Mizuho will study the future construction of a GPU cluster^{*3} connecting multiple GPU servers. As a computing platform capable of supporting production use, Mizuho will examine the required architecture and operational model from the perspectives of AI workloads including inference, networking, storage, LLMOps, security, operations management, and scalability.

Through these efforts, Mizuho aims to realize an AI infrastructure that appropriately combines cloud and on-premises environments, enabling flexible use of diverse computing resources according to use case, confidentiality, performance, and cost.

^{*1} NVIDIA DGX B200: A high-performance computing platform that supports AI workloads from model development to production operations. Equipped with eight NVIDIA Blackwell GPUs and connected via fifth-generation NVIDIA NVLink, it enables efficient training and inference of large-scale LLMs.

^{*2} Mizuho LLM: A large language model being developed by Mizuho with the aim of enabling responses that

reflect the specialized knowledge and internal rules required for financial operations.

Japanese press release regarding Mizuho LLM:

<https://prtimes.jp/main/html/rd/p/000000002.000177612.html>

*3 GPU cluster: A computing environment in which multiple GPU servers are interconnected to efficiently perform large-scale LLM training, evaluation, and inference.

2. Technical validation of a secure AI agent execution environment using NVIDIA NemoClaw

Mizuho will also conduct technical validation of an execution environment for the secure use of AI agents at financial institutions by leveraging NVIDIA NemoClaw^{*4}.

In this validation, Mizuho will examine key issues including execution environment isolation, data protection, network access control, permission management, and verification of execution history, assuming that AI agents are integrated with internal data and business systems. In parallel, Mizuho will also validate integration with generative AI models including the Mizuho LLM currently under development, and assess the potential of AI agent use based on internal business context.

Through these validations, Mizuho aims to realize the use of AI agents in areas involving highly confidential data and internal systems while balancing security and scalability.

*4 NVIDIA NemoClaw: A collection of reference blueprints for securely executing and managing AI agents built using NVIDIA Nemotron open models, open harnesses, and the NVIDIA OpenShell secure runtime. Through execution environment isolation and network access control enabled by OpenShell, it supports the use of AI agents within enterprises with due consideration for security and governance requirements.

Outlook

Going forward, Mizuho aims to securely integrate generative AI models, including the Mizuho LLM currently under development, with highly confidential data and internal systems, thereby expanding the scope of AI agent utilization while ensuring the governance and control required by financial institutions.

Through this effort, Mizuho will continue to build an environment in which employees can use AI agents as partners in their day-to-day work, enabling more efficient information gathering, document preparation, analysis, and development support. By expanding each employee's expertise and enabling them to devote more time to dialogue with customers and higher value-

added activities, Mizuho seeks to enhance the value it delivers to customers.

Mizuho positions AI not merely as a tool for efficiency, but as an important management foundation for enhancing the expertise of each employee and increasing customer value. Going forward, through collaboration with advanced domestic and global technology partners, Mizuho aims to create a model case for secure and trusted AI utilization at financial institutions.

End

About Mizuho Financial Group, Inc.

Company Name: Mizuho Financial Group, Inc.

Representative: Masahiro Kihara, President & Group CEO

Head Office: Otemachi Tower, 1-5-5 Otemachi, Chiyoda-ku, Tokyo 100-8176, Japan