



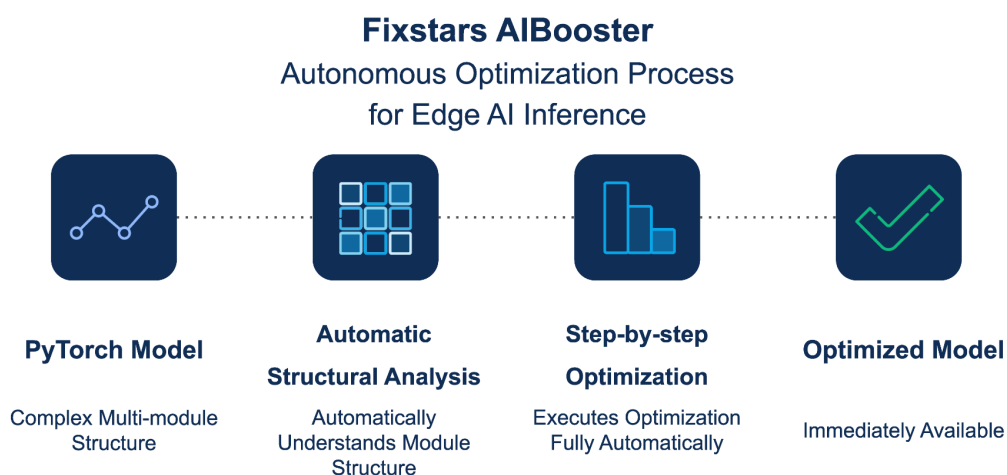
FOR IMMEDIATE RELEASE

Fixstars Adds Autonomous Optimization Feature for Edge AI Inference to its AI Acceleration Platform "AIBooster"

Enhances Autonomous Optimization for AI Training and Achieves Dramatic Improvement in AI Development Efficiency; Introduces SaaS Plan

IRVINE, California – Oct 09, 2025 – Fixstars Corporation (TSE Prime: 3687, US Headquarters: Irvine, CA), a leading company in performance engineering technology, today announced the release of the latest version of its AI processing acceleration platform, "Fixstars AIBooster" (hereinafter "AIBooster"), with significantly expanded performance and features.

The latest version of AIBooster features two new functionalities: autonomous optimization via model compression for Edge AI inference, and autonomous hyperparameter optimization for AI training. Furthermore, a new plan offering the performance observation feature as a Software-as-a-Service (SaaS) solution has been introduced to significantly reduce the effort required for initial deployment and allow immediate use. Fixstars will continue to support customers in making their AI development cycles more efficient and faster, enabling AI utilization across various devices.

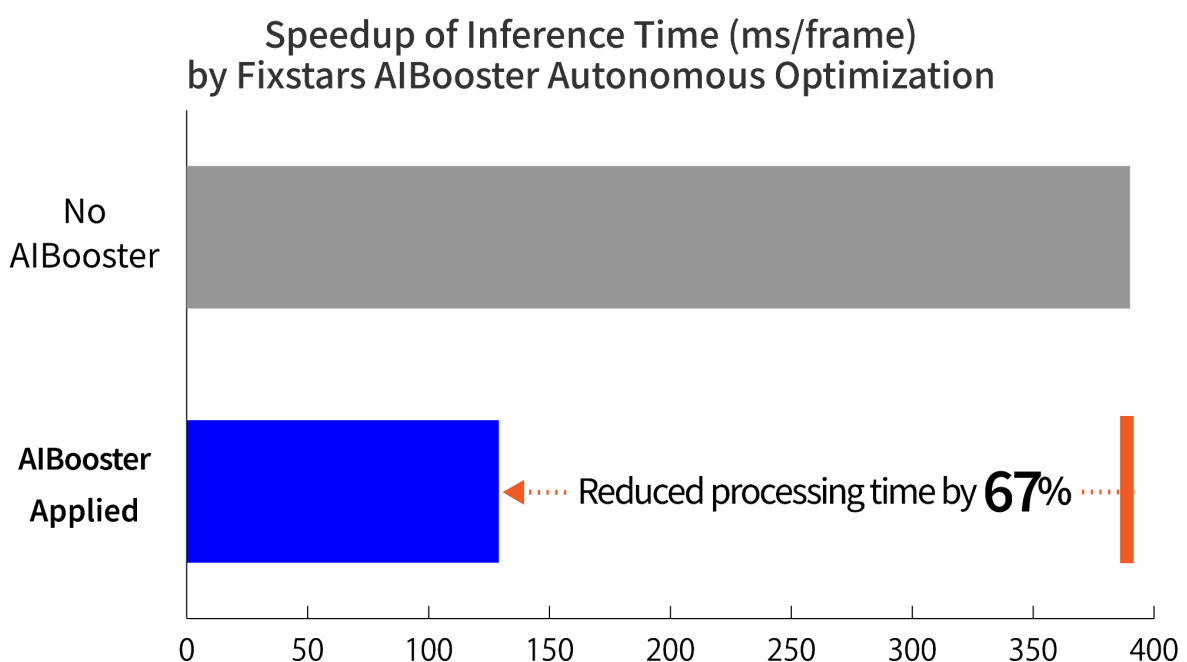


New Features of the Latest AIBooster

1. Autonomous Optimization for Edge AI Inference

When performing AI inference on local devices such as in-vehicle systems or mobile devices (Edge AI inference), constraints on computing resources and memory capacity make it difficult to operate AI models as is. This previously required a great deal of trial-and-error to tune the AI model for optimal performance on each device. AIBooster's new autonomous optimization feature addresses this by automatically converting the AI model for edge devices while maintaining performance.

This function targets AI models created using the common framework PyTorch and automatically converts them to a format that maximizes performance on each device, utilizing a model conversion backend. The initial release supports NVIDIA TensorRT, optimizing inference processing in NVIDIA GPU environments. This capability not only dramatically improves inference speed through techniques like quantization and kernel fusion but also significantly reduces development man-hours by automating the conversion process.



2. Enhanced Hyperparameter Autonomous Optimization for AI Training

In the AI model training process, hyperparameter settings, such as batch size and learning rate, greatly influence model accuracy and training efficiency, but their optimization requires significant time and trial-and-error. The newly added hyperparameter autonomous optimization for AI training integrates optimization across the model, resources, and hardware to maximize overall training performance.

Key features include:

- **Integrated Optimization:** Balances model accuracy and training efficiency while maximizing resource utilization.
- **Hardware Control:** Automatically adjusts CPU core and GPU usage schedules.
- **Distributed Learning Support:** Operates efficiently in large-scale environments such as Slurm and Kubernetes.

The system outputs trial data for optimization, along with initial and optimized settings and improvement rates for review. By visualizing and automating the trial-and-error optimization process, the feature significantly contributes to improving AI training efficiency and acquiring knowledge.

3. SaaS-based Performance Observation Function

The latest AIBooster introduces a SaaS-based performance observation function that eliminates setup and operational hassle, allowing users to start immediately.

The SaaS version saves performance data collected from GPU servers to the cloud and visualizes it in real-time on a dashboard. It supports multi-cloud environments and server groups distributed across multiple locations, allowing users to grasp the status of the entire system, down to node-level and computation job-level details, all on one screen. Fixstars is responsible for managing the security of the SaaS environment, ensuring the safe protection of customers' AI development information. Support for on-premise installation is also continued, offering flexible usage tailored to specific needs and operational policies.

Related Seminar: "NVIDIA Latest Embedded GPU × Fixstars AIBooster — Fastest Path to Edge AI Inference Operations"

A co-hosted seminar with NVIDIA Japan will be held to cover the features of the latest AIBooster, including its autonomous optimization for Edge AI inference, as well as the performance and utilization methods of the latest GPUs.

The poster is for a seminar titled "NVIDIA最新組み込みGPU × Fixstars AIBooster エッジAI推論を最速で実運用へ" (NVIDIA Latest Embedded GPU × Fixstars AIBooster: Accelerating Edge AI Inference to Real-world Operation). It features three speakers: 國本 純希 (Junaki Kunitani), 中根正雄 (Masao Nakane), and 孫正道 (Masahito Sonoda). The event is scheduled for 10/23 (木) 12:00~13:00 and is online with no fee. Logos for Fixstars and NVIDIA are at the top.

Speed up your AI
FIXSTARS

NVIDIA

NVIDIA最新組み込みGPU × Fixstars AIBooster

**エッジAI推論を
最速で実運用へ**

10/23 (木) 12:00~13:00

オンライン 参加費無料

國本 純希
(株) フィックスターズ
ソリューション第二事業部 シニアエンジニア

中根正雄
Business Development Manager
Enterprise, NVIDIA

孫正道
(株) フィックスターズ
モビリティ事業部 事業部長

Format: Online (Zoom)

Date/Time: Thursday, October 23, 2025, 12:00 PM – 1:00 PM JST

Organizers: Fixstars Corporation, NVIDIA Japan

Title:

NVIDIA Latest Embedded GPU × Fixstars AIBooster “Fastest path to Edge AI Inference Operations”

Registration: <https://news.fixstars.com/5749/>

###

About Fixstars Corporation

Fixstars is a technology company dedicated to accelerating AI inference and training through advanced software optimization solutions. It supports innovation in healthcare, manufacturing, finance, mobility, and other industries. For more information, visit: <https://www.fixstars.com/>

Media Contact

Public Relations, Fixstars Corporation

Email: press@fixstars.com

Tel: +81-3-6420-0751