Forward Deployed Engineer (AI Infrastructure Engineer)

worklife balance, Advance Your AI Career

Job Information

Hiring Company
systemsGo Corporation

Subsidiary
systemsGo

Job ID
1615963

Industry
IT Consulting

Company Type
Small/Medium Company (300 employees or less) - International Company

Non-Japanese Ratio
About half Japanese

Job Type
Permanent Full-time

Location
Tokyo - 23 Wards

Salary
Negotiable, based on experience

Work Hours
Mon-Fri 9am-6pm

Holidays
Starts at 13 days/yr paid leave, increases each year until 22/yr

Refreshed
October 19th, 2026 00:00

General Requirements

Minimum Experience Level
Over 3 years

Career Level
Mid Career

Minimum English Level
Business Level (Amount Used: English usage about 50%)

Minimum Japanese Level
Business Level

Minimum Education Level
Technical/Vocational College

Visa Status
Permission to work in Japan required

Job Description

About the Role

Our client is a global technology company developing next-generation AI infrastructure solutions. They are seeking a Forward Deployed Engineer to support the deployment, integration, and operation of large-scale AI systems within customer environments across Japan.

Key Responsibilities

- Deploy and integrate AI infrastructure solutions in customer environments.
- Manage and optimize Kubernetes-based platforms and Helm deployments.
- Troubleshoot issues across AI platforms, distributed systems, networking, and infrastructure.
- Develop automation tools, scripts, and operational workflows.
- Collaborate closely with customers and internal engineering teams to ensure successful deployments and ongoing operations.
- Serve as a trusted technical advisor throughout implementation and production support.

Required Skills

Required Qualifications

- 5+ years of experience in Infrastructure Engineering, Platform Engineering, SRE, MLOps, Machine Learning Engineering, or related roles.
- Strong software engineering skills with advanced Python proficiency.
- Experience with Linux, Kubernetes, Helm, CI/CD, and infrastructure automation.
- Solid understanding of distributed systems and cloud-native technologies.
- Experience supporting production environments and troubleshooting complex systems.
- Fluent Japanese and professional-level English communication skills.

Preferred Qualifications

- Experience with AI/ML infrastructure, HPC, or large-scale compute environments.
- Knowledge of AI serving frameworks such as vLLM or SGLang.
- Experience with accelerator-based computing (GPU/AI hardware).
- C++ development experience.
- Enterprise customer-facing experience.

Language

- Japanese: Business level
- English: Business level

Company Description