

# G Talent

Professional IT Recruitment | 外国人ITエンジニアの転職支援  
 - Many IT jobs for global companies and high-profile startups!  
 - IT specialty bilingual consultants support your career change!

## MLOps Engineer

**Remote Work★Global Business**

### Job Information

#### Recruiter

G Talent at Bizmates, Inc.

#### Hiring Company

★Multi-core processor-related company★

#### Job ID

1484139

#### Industry

Internet, Web Services

#### Job Type

Permanent Full-time

#### Location

Tokyo - 23 Wards

#### Salary

5.5 million yen ~ 8.5 million yen

#### Work Hours

Discretionary Work System

#### Holidays

Saturday/Sunday/National Holiday, Annual Paid Leave, etc.

#### Refreshed

July 5th, 2024 07:00

### General Requirements

#### Minimum Experience Level

Over 3 years

#### Career Level

Mid Career

#### Minimum English Level

None

#### Minimum Japanese Level

Business Level

#### Minimum Education Level

High-School or Below

#### Visa Status

Permission to work in Japan required

### Job Description

#### 【About the company ...】

The company is a software company whose corporate message is "Speed up your Business."

They accelerate their customers' businesses in various fields such as medical, manufacturing, finance, and entertainment by parallelizing and optimizing software to efficiently utilize multi-core processors and accelerating applications using new

memory technology that realizes power saving and high-speed IO, thereby achieving green IT.

## **[ Job Description ]**

### **[Overview]**

They are developing machine learning infrastructures for efficient development of high-performance machine learning models.

Specifically, they are researching and developing lightweight methods such as quantization and branch pruning of deep learning models, domain-specific compilers that automatically speed up the inference process of models, and HW aware NAS infrastructure that can be recognized efficiently on the target hardware, and building MLOps using these methods.

### **[Responsibilities]**

- Development of deep learning compilers
- Research and development of lightweight methods for deep learning models such as quantization and branch pruning
- Development of HW aware NAS
- Construction and operation of MLOps using the above

### **[Rewarding]**

- Challenge new technological innovations in the fast-growing AI market
- Contribute to the development of well-known OSS
- Experience building systems using public cloud services
- Surrounded by top-notch engineers, improve your skills through daily work and discussions

### **[Development Environment]**

Development environment: C, C++, Python3, Bash

Framework : TensorFlow

Other development environment: Linux, Debian

Development Support Tools: JIRA, Git, GitLab, Apache Airflow

Development Methodology: Agile, Scrum, Global team (multinational members)

Development content type: B2B, machine learning, AI, mobility-related (self-driving, traffic-related)

Cloud platform: Amazon Web Service

Infrastructure management: Docker, Terraform, Ansible, Kubernetes, Amazon ECS, Amazon Elastic Kubernetes Service

AI and data analysis: Apache Beam, Keras, PyTorch, pandas, Jupyter notebook, matplotlib, NumPy

---

## **[ Working time ]**

Discretionary Work System

### **[ Welfare ]**

- Full social insurance
- Commuting allowance
- Child allowance
- Recreation Assistance
- Training/Self Development Allowance Resident allowance of 30,000 yen/month
- Travel and accommodation expenses for company trips
- Company expenses for social gatherings
- Maintenance fee for club activities (e.g., ski club: 60,000 yen per person per year; other clubs include golf, futsal, and board game clubs)

### **[ Holiday ]**

- 2 days a week (Saturday and Sunday)
- National Holidays
- Summer vacation
- New Year holiday
- Annual Paid leave, etc.

---

## **Required Skills**

### **[Required]**

- Experience in development using C/C++ and Python
- Basic knowledge of deep learning

### **[Preferred]**

- Experience using or developing deep learning frameworks and deep learning model formats such as TensorFlow, PyTorch, ONNX, etc.
- Experience developing domain-specific compilers such as TVM, Halide, MLIR, etc.
- Experience in software acceleration of deep learning inference processes
- Knowledge or development experience with deep learning accelerators
- Development experience using cloud infrastructure such as Amazon Web Service or Azure
- Experience in building and operating MLOps

---

Company Description