

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!

【Dev】 Machine Learning Engineer (LLM New Business)

Flextime★FullyRemoteWork

Job Information

Recruiter

G Talent at Bizmates, Inc.

Hiring Company

※AI/NLP Company※

Job ID

1517242

Industry

Internet, Web Services

Job Type

Permanent Full-time

Location

Tokyo - 23 Wards

Salary

7 million yen ~ 12 million yen

Work Hours

Flextime System

Refreshed

January 24th, 2025 17:15

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

Visa Status

No permission to work in Japan required

Job Description

[About the Company]

The company's mission is "to advance humanity by reinventing the mechanisms of value creation" and aims to create an environment where people can focus on creating new value through the use of natural language processing and AI technologies.

AI delivers information directly related to business operations, thereby enhancing the information sensitivity of the organization. In addition, the company offers two SaaS products: "Anews," which encourages the conception of business ideas and their development within the organization, and "Astrategy," a market research service that provides suggestions by

structuring necessary information according to the organization and its operations.

Currently, its main target is major manufacturing R&D in Japan, and more than 30% of the Nikkei 225 companies have already adopted the system. In the future, starting with manufacturing R&D, the company plans to expand to domestic enterprise companies, then to Japanese companies as a whole, and even to global companies.

The company completed a Series C financing of 1.1 billion in August 2022.

In February 2024, they were selected for the "GENIAC (Generative AI Accelerator Challenge) Project" promoted by the Ministry of Economy, Trade and Industry and NEDO, and have begun to develop their own LLM with 100 billion parameters! The company has grown to 100 employees and is looking for more personnel to achieve further growth!

[Job Description]

[Company Overview]

The company's mission is "to advance humanity by reinventing the mechanism of value creation," and it supports the corporate transformation of many companies by utilizing the most advanced generative AI technology.

The company operates "Anews," a one-stop search engine for internal and external information, and "SAT (Stockmark A Technology)," a system for structuring all types of data and turning them into corporate assets. We also support the development of company-specific generative AI and the construction of proprietary systems.

Currently, the main target is major manufacturing R&D in Japan, which already accounts for 30% of the Nikkei 225 and has been introduced in over 250 companies. In the future, starting with manufacturing R&D, the company plans to expand to domestic enterprise companies, then to Japanese companies as a whole, and even to global companies.

The company is also focusing on research and development of a purely domestic LLM that is strong in the Japanese language and business domain, and in May 2024, the company released its own LLM with 100 billion parameters!

In October 2024, it completed a Series D fundraising of 4.5 billion yen, for a total of 8.8 billion yen.

With over 100 employees, the company is recruiting personnel to achieve further growth!

[Current Issues]

They have launched a new business in 2024 and have launched several PoC projects for the enterprise. The team is working hard to ensure the success of the projects, but we are still facing several challenges.

- Lack of research and development resources to improve the accuracy and expand the functionality of the document analysis engine.
- Lack of research and development resources to expand the functionality of the RAG verification service (in-house SaaS)

[Your Role]

As a machine learning engineer for a new SaaS business using LLM, you will lead the planning, verification, and execution of product improvement plans using machine learning.

In the execution of measures, the position involves a lot of collaboration with internal product engineers, PdMs, and researchers.

■Responsibilities

- Improve the accuracy of the system for structuring various documents such as HTML/PDF/PowerPoint using natural language processing and image processing.
- Build, train, and evaluate training datasets for ViT-based models and VLMs.
- Verification and implementation of query expansion, Chunking, Reranking, and answer generation algorithms.

Scope of change: Development related work

[Team Structure]

PaaS Unit

1 Business Manager (CEO)

1 Product Manager

[Development Environment]

[Organizational Structure]

- 1 Product Manager
- 1 front-end engineer
- 1 backend engineer
- 3 machine learning engineers

[Development Languages]

TypeScript, Vue.js, Node.js, Python

[Containers]

Docker

[IaC]

Terraform

[Cloud]

AWS, Azure

[Attractiveness of this position]

- You can try to service business document analysis that has not been achieved by general-purpose LLM or major cloud-based document analysis services.
- You can experience LLM/multi-modal AI as a real-world service.

[Working conditions & treatment]

- Salary increase (twice a year / May and November)

[About remote work]

- Basically, fully remote work
Residence must be in Japan (no full remote work while residing overseas)
- Flexible remote/working options (no restrictions on the number of times you can come to the office)
Each employee works while using the system according to his/her own circumstances.

There are people who work full remotely from Sendai, Kyoto, Osaka, etc., people who enjoy working while on vacation, foreign nationals who return to their home countries for a few weeks to work, and people who mainly come to work, making for a diverse work style.

[Welfare]

- Customer facing allowance (flat monthly amount of 20,000 yen *included in monthly salary)
Free use for self-improvement and customer value improvement
(purchase of books/attendance at seminars/equipment, etc.)
- Lump-sum allowance for preparation for remote work (flat rate of 50,000 yen in the first paycheck)
Provided as an expense for preparation for remote work

[Others]

- Periodic health checkups (all paid by company)
- Influenza vaccination
- PC of your choice provided
- Free drinks such as water/coffee/carbonated drinks/alcohol provided at the office
- Office Glico installed

[Education and Skill Development]

- Onboarding for 5 days after employment is available
- Monthly study day system (for own skill improvement and learning purpose)

[Internal Communication]

- Weekly company-wide online meeting
- Monthly social gathering (offline, some online)
- Semi-annual kick-off (hybrid of online and offline meetings)

[Holidays and Vacations]

- 123 days off per year
- Saturdays, Sundays and national holidays (2 days off per week)
- Year-end and New Year vacations
- Paid leave (granted at the time of employment *Number of days granted depends on the month of employment)
- Prenatal leave
- Childcare leave

Required Skills**[Required]**

- *Meet all of the following
 - Completion of a graduate degree in science or engineering
 - 3+ years of experience as an engineer or researcher working with machine learning
 - Experience developing services using Python

[Preferred]

- Master's or PhD in a technical field related to computer science
- Experience in development/operation using machine learning in in-house services for customers
- Experience with machine learning models for text and images, VLMs, or research experience
- Experience developing in the cloud
- Experience in a leadership role in understanding the business background and planning/executing short-term measures on your own

[Ideal Applicants]

- Love computer science and distributed processing
 - Love to follow and verify the latest papers and case studies
 - Able to ask questions on your own and create solutions to the problems you are given.
 - Able to communicate with the business side and think about how to contribute to the product.
 - Positive spirit to take on challenges
-
- Able to live in Japan
 - Able to communicate smoothly in Japanese, including technical details related to development.
Guideline: Daily conversational level

Company Description