

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】 Back-end Engineer (LLM New Business)

Flextime★FullyRemoteWork

Job Information

Recruiter

G Talent at Bizmates, Inc.

Hiring Company

※AI/NLP Company※

Job ID

1522513

Industry

Internet, Web Services

Job Type

Permanent Full-time

Location

Tokyo - 23 Wards

Salary

7.5 million yen ~ 12 million yen

Work Hours

Flextime System

Refreshed

April 1st, 2025 00:00

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

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]

[About the company]

With the mission of "Reinventing the Mechanism of Value Creation to Advance Humanity," the company aims to create an environment where people can focus on creating new value through the use of natural language processing and AI technology.

AI delivers information directly related to business operations, thereby increasing the information sensitivity of the organization. The company also offers two SaaS products: "Anews," which encourages the conception of business ideas and their development within the organization, and "Astrategy," a market research service in which AI structures the necessary information according to the organization and its operations and provides suggestions, thereby supporting the creation of new value for companies.

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.

Completed Series C financing of 1.1 billion in August 2022.

In February 2024, the company was selected for the "GENIAC (Generative AI Accelerator Challenge) Project" promoted by the Ministry of Economy, Trade and Industry and NEDO, and in May 2024, the company released its own LLM of 100 billion parameters!

They have more than 100 employees and are recruiting personnel to achieve further growth!

[Issues]

In 2024, the company launched a new business and several enterprise PoC projects have been launched and the team is struggling to make the projects successful, but is still facing multiple challenges.

- Verify, develop, and operate the back-end and cloud infrastructure for a new service (PaaS) that provides a document analysis engine.
- Defining new resource requirements and interfaces

[Responsibilities]

You will be responsible for the back-end to infrastructure areas of Stockmark A Technology, a new business that promotes the use of generative AI.

- Validation, implementation, and product operation of the business document analysis engine API and its infrastructure
 - Design, construction, and operation of overall infrastructure (RAG data pipeline, customer data management) in Azure environment
 - Design, build, and product operation of data flow and networking, including customer environments
- All of this is both PoC and Production

*Scope of change: development-related tasks

[Team Structure]

PaaS Unit
1 Business Manager (CEO)
1 Product Manager

[Development Environment]

[Organization]

- 1 Product Manager
- 1 MLOps engineer
- 2 Machine Learning Engineers

[Languages/frameworks used]

TypeScript, Vue.js, Node.js, Python

[Containers]

Docker

[IaC]

Terraform

[Cloud]

AWS, Azure

[Attractiveness of this position]

- You will be the first backend engineer in a new business and will be deeply involved in decision making, such as technology selection and architecture.
- You can be involved in the full cycle of product development using generative AI, from design to operation.
- You can experience the cutting-edge LLM technology before anyone else in the world and implement it in society.

· You will be able to work together with the different talents of the Tech organization, mainly PdMs, designers, and engineers who have been active at the forefront of their respective industries.

[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

- Completed a graduate degree in science or engineering
- 5+ years of experience in engineering
- Experience developing/operating services using AWS, Azure

[Preferred]

- Master's or PhD in a technical field related to computer science
- Experience developing/operating in-house developed services for customers
- Experience in machine learning experiments, model management, data management, and building learning and inference pipelines using AWS
- Experience in development/operation using Terraform
- Experience in understanding the business background and planning/executing measures on your own in a leadership position

[Ideal Applicants]

- Love computer science
 - Love to follow and test the latest theories and case studies
 - Able to ask questions on your own and generate solutions for them, not just for a given problem
 - Able to communicate with the business side and think about ways to contribute to the product
 - Positive spirit to take on challenges
- Residence in Japan is required.
 - Able to communicate smoothly in Japanese, including technical details related to development.
 - Reference: Daily conversational level

Company Description