MLOps Engineer @ The renewable energy company

English speaking only available

Job Information

Recruiter
PROGRE Ltd

Hiring Company
The renewable energy company

Job ID
1625181

Industry
Internet, Web Services

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

Non-Japanese Ratio
About half Japanese

Job Type
Permanent Full-time

Location
Tokyo - 23 Wards, Chiyoda-ku

Salary
8 million yen ~ 15 million yen

Work Hours
Super Flex Time

Refreshed
November 27th, 2026 13:28

General Requirements

Minimum Experience Level
Over 3 years

Career Level
Mid Career

Minimum English Level
Business Level (Amount Used: English Only)

Minimum Japanese Level
None

Minimum Education Level
Bachelor's Degree

Visa Status
Permission to work in Japan required

Job Description

As an MLOps Engineer, you will own significant parts of this lifecycle and serve as the engineering bridge between research code and reliable production systems. This is a platform-and-operations engineering role, rather than a role focused on conducting model research.

The team is developing AI-native pipelines that use AI and LLMs to automate the ML lifecycle—from operations and monitoring through deployment—and a "digital twin / time machine" validation platform that can reproduce and test historical states at arbitrary points in time. We are looking for someone who is interested in power and market data, takes initiative, and turns operational problems into robust systems.

Key Responsibilities

- Design and build AI-native pipelines that use AI agents and LLMs to automate ML operations end to end, including anomaly monitoring and first-line triage, retraining and deployment decisions, and operational documentation.
- Design and build a digital twin / time machine validation platform, including point-in-time reproducible backtesting using as-of and versioned data, and pre-release shadow, replay, and scenario testing.
- Work with data scientists to design experiments and turn model prototypes into reproducible, testable, and deployable workflows.
- Design, build, and operate feature-engineering pipelines for batch and, where necessary, near-real-time training and prediction.
- Build model testing and validation frameworks covering data requirements, training behavior, prediction outputs, and deployment gates.
- Implement evaluation and monitoring for prediction quality, data drift, concept drift, feature skew, latency, and pipeline health.
- Design automated workflows for model training, retraining, deployment, rollback, and version management.
- Help scale the platform to support new models and forecasting use cases; identify and communicate architectural trade-offs and implementation risks.
- Improve the reliability, security, cost efficiency, and observability of the ML platform, and respond to production incidents.
- Collaborate with Data, Platform, and SRE teams on shared data, orchestration, infrastructure, and governance foundations.
- Identify problems independently and propose and drive improvements in collaboration with data scientists, business teams, and other engineering groups.

Technology Environment

Go, Python, SQL, REST, gRPC, Protocol Buffers, FastAPI, Gin, Echo, GCP (Cloud Run, Cloud Workflows, Cloud Scheduler, Pub/Sub), relational databases, ORMs, Docker, Terraform, GitHub, GitHub Actions, and Cloud Build.

The role focuses on software systems for trading, balancing, planning, and market integration. Development of machine-learning models themselves is outside the scope of this position.

Required Skills

Required Qualifications

- Experience building AI-native ML pipelines: using AI agents, LLM-driven automation, or AI coding tools to automate ML processes from operations and monitoring through model deployment. You should also use tools such as Cursor, Claude, or ChatGPT in your day-to-day work. Experience limited to introducing or operating conventional MLOps tools is not sufficient.
- Approximately 4+ years of relevant experience in data science, ML engineering, MLOps, data engineering, or a related production role, with hands-on ownership of production ML lifecycle components using Python. This includes model deployment, training and prediction pipelines, monitoring, model versioning, and ML CI/CD.
- Experience designing or building a digital twin / time machine-style validation capability. Relevant examples include point-in-time reproducible backtesting using as-of or versioned data, preventing look-ahead bias and data leakage, shadow or replay testing, scenario simulation, or efficient experimentation and validation workflows.
- A track record of identifying problems independently and initiating improvements, such as building new systems or tools, introducing technology, or improving development and operational processes.
- Strong communication skills and business-level English, including the ability to collaborate with data scientists, business stakeholders, and engineering teams, and to produce engineering documentation in English.

Preferred Qualifications

- Energy-related modeling experience, such as weather forecasting, renewable generation, electricity demand, commodity prices, or other time-series problems.
- Experience with time-dependent modeling and the deployment requirements of real-time and batch feature, training, and prediction pipelines.
- Experience building or operating LLMOps or AI-agent platforms, including evaluation, guardrails, tracing, and cost management.
- Hands-on experience with ML/data orchestration tools such as Airflow, Metaflow, Kubeflow, Vertex AI Pipelines, Dagster, or Prefect.
- Experience designing model testing frameworks, model evaluation, retraining routines, and deployment controls.
- Experience building monitoring platforms and dashboards using Grafana, Looker, Metabase, or similar tools.
- Experience with Vertex AI, MLflow, Feast, model registries, feature stores, or experiment-tracking platforms.
- Practical experience with GCP, Docker, and infrastructure-as-code tools such as Terraform, Terragrunt, or Ansible.
- Experience with automated testing, Git-based development, CI/CD for ML applications, and progressive delivery.
- Experience with data quality frameworks, data lineage, governance, and reproducibility controls.
- Experience with Kubernetes, distributed processing, GPU workloads, or performance and cost optimization.
- Experience in observability and incident response for production ML systems.
- Experience in the energy, electricity trading, or fintech sectors.
- Experience working in startups, Agile environments, or multicultural teams.
- Japanese language skills for business and domain communication.

Company Description