



シンガポールの求人なら
JAC Recruitment Singapore

PR/097416 | Machine Learning (Engineer | Lead)

Job Information

Recruiter

JAC Recruitment Singapore

Job ID

1604812

Industry

IT Consulting

Job Type

Permanent Full-time

Location

Singapore

Salary

Negotiable, based on experience

Refreshed

July 28th, 2026 10:26

General Requirements

Minimum Experience Level

Over 3 years

Career Level

Mid Career

Minimum English Level

Fluent

Minimum Japanese Level

None

Minimum Education Level

Associate Degree/Diploma

Visa Status

No permission to work in Japan required

Job Description

What You'll Be Responsible For End-to-End ML Systems Ownership

Lead the design, implementation, and operation of the complete machine learning lifecycle, including data pipelines, training workflows, evaluation frameworks, inference infrastructure, deployment processes, and production monitoring.

Model Adaptation & Optimization

Fine-tune and optimize large language models and foundation models using modern techniques such as LoRA, QLoRA, Supervised Fine-Tuning (SFT), Direct Preference Optimization (DPO), distillation, and other emerging approaches to improve quality, efficiency, and task performance.

Scalable Inference Architecture

Design and maintain robust inference systems capable of serving production workloads while balancing latency, throughput, reliability, and infrastructure cost.

Data Systems & Training Infrastructure

Develop and manage data pipelines that support the collection, generation, validation, and maintenance of high-quality datasets, leveraging both synthetic and real-world data sources to continuously improve model performance.

Evaluation & Quality Assurance

Establish comprehensive evaluation methodologies to measure model accuracy, robustness, safety, bias, reliability, and user-impact metrics. Work closely with stakeholders to ensure quality standards are clearly defined and consistently met.

Production Readiness & Performance Engineering

Own production deployment and operational excellence, including GPU utilization, memory optimization, model serving efficiency, performance tuning, observability, reliability engineering, and scaling strategies.

Cross-Functional Collaboration

Partner closely with product engineers, backend teams, platform engineers, and other stakeholders to integrate machine learning capabilities seamlessly into user-facing products and workflows.

Technical Leadership

Provide technical direction, make sound engineering decisions, and help establish best practices that enable the team to move quickly while maintaining high standards of quality and reliability.

Continuous Improvement

Drive iterative improvements by leveraging production feedback, operational metrics, user insights, and experimentation to enhance system performance and overall user experience.

What Success Looks Like

In this role, you will:

- Consistently transform machine learning research, prototypes, and experimentation into reliable production solutions.
- Build ML infrastructure that is scalable, maintainable, and easy to operate.
- Establish efficient training, evaluation, and deployment workflows that accelerate iteration without compromising quality.
- Proactively identify and resolve production issues before they impact users.
- Enable the broader team to work effectively through strong technical leadership, clear communication, and collaborative problem-solving.
- Deliver measurable improvements in model quality, system reliability, operational efficiency, and user outcomes.

Technology Environment

You will work extensively with:

- Python
- PyTorch and/or JAX
- GPU-accelerated training and inference platforms
- Distributed machine learning systems
- Model serving and deployment infrastructure
- Modern data processing and orchestration frameworks

What We're Looking For

- Proven track record of building, deploying, and maintaining machine learning systems in production environments.
- Strong understanding of large language models, foundation models, and their practical limitations, trade-offs, and failure modes.
- Experience developing scalable ML infrastructure, training pipelines, and inference systems.
- Strong software engineering fundamentals with a focus on reliability, maintainability, and performance.
- Demonstrated ability to take ownership of complex technical initiatives from concept through production deployment.

Preferred Attributes

- Comfortable operating in fast-moving environments where priorities evolve quickly and execution matters.
- Pragmatic problem solver who balances technical excellence with business impact.
- Strong communicator capable of collaborating across technical and non-technical teams.
- Passion for building products that deliver meaningful value to users at scale.
- Natural leadership qualities with the ability to align teams around technical goals and outcomes.

Our Working Style

We believe exceptional products are built by small, highly capable teams with a strong sense of ownership and accountability. We value thoughtful decision-making, rapid learning, and a bias toward execution.

Team members are encouraged to contribute ideas, challenge assumptions, and independently drive initiatives forward. We strive for a culture where technical excellence, collaboration, and product impact are equally important.

Jaspreet Kaur Sran (R22109724)

JAC Recruitment Pte. Ltd. (90C3026)

#LI-JACSG

Notice: By submitting an application for this position, you acknowledge and consent to the disclosure of your personal information to the Privacy Policy and Terms and Conditions, for the purpose of recruitment and candidate evaluation.

Privacy Policy Link: <https://www.jac-recruitment.sg/privacy-policy>

Terms and Conditions Link: <https://www.jac-recruitment.sg/terms-of-use>

Company Description