

Prashant Anand

Staff Machine Learning Engineer

Staff ML Engineer with 6 years at Mercari (Tokyo) building production ML systems for Japan's largest C2C marketplace. Lead ML/NLP/LLM applications on Japanese text in Customer Support and Trust & Safety.

Currently leading ML-driven product moderation system for Mercari Shops. Shipped automatic high-risk inquiry escalation, LLM-powered auto-replies, and intelligent inquiry routing.

B.Tech, IIT Delhi (2019). Speaker, PyCon JP 2024 and PyCon APAC 2023.

- Tokyo, Japan
- @primaprashant
- in/primaprashant
- anandprashant.com
- heyanandprashant@gmail.com

EDUCATION

Indian Institute of Technology Delhi
B.Tech. in Electrical Engineering
New Delhi, India • Jul 2015 – May 2019

PUBLICATIONS

- Few Shot Speaker Recognition using Deep Neural Networks
- Data Driven Sensing for Action Recognition using Deep Convolutional Neural Networks

SKILLS

SWE

- Python
- Kubernetes
- Docker
- GCP
- BigQuery
- gRPC
- MySQL
- Terraform

ML

- PyTorch
- Transformers
- OpenAI API
- TorchServe
- Datasets
- Vertex AI
- W&B

AI Tools

- Claude Code
- Cline
- Open WebUI
- ChatGPT

WORK EXPERIENCE

Machine Learning Engineer Oct 2019 – Present
Mercari, Inc. • Tokyo, Japan

Progression: MG1 (10/2019) → MG2 (1/2021) → MG3 (7/2021) → MG4 (8/2024)

- Leading the design and implementation of the ML-driven product moderation system for B2C marketplace, Mercari Shops, defining the technical roadmap and coordinating with product and TnS teams.
- Designed and deployed a real-time high-risk inquiry detection system using a stacked ensemble of fine-tuned ModernBERT and GPT-4.1, improving detection by **51%** and cutting escalation time by **35%**.
- Shipped a DistilBERT-based auto-reply system that handles **13%** of total inquiries, saving **2.200+** agent hours monthly. The system integrates 16 metadata features from 7 microservices.
- Launched a GPT-3.5 auto-reply system for low-complexity inquiries, reducing manual replies by **39%** and agent handling time by **6%** within its target segment.
- Optimized inquiry routing by productionizing fine-tuned Japanese DistilBERT model on 40+ skill categories, increasing assignment accuracy by **8%** and reducing inter-agent transfers by **26%**.

TALKS

Tokyo AI, AMA #09 Jun, 2025
Intuitive Jargon-Free Mental Model for MCP

PyCon JP 2024 Sep, 2024
AI-powered Automatic Replies in Customer Support (Slides)

PyCon APAC 2023 Oct, 2023
AI for Efficient Routing of Customer Inquiries at Mercari (Slides)