

Tanaka Yuma

Senior Software Engineer

tanakayuma90@outlook.com | (567) 396-3172 | Tokyo, JP

Professional Summary

Based in Tokyo, I've spent 12 years building at the intersection of AI/ML and full-stack engineering. From deploying large language models in production to designing real-time inference pipelines and intuitive frontend experiences, I'm passionate about shipping AI products end-to-end.

Skills

- **Languages:** Java, Go, Python, C/C++, C#, Kotlin, Ruby, JavaScript, TypeScript, SQL, T-SQL, HTML5, CSS3, Bash, Swift
- **Frontend:** React, React Native, Angular, Next.js, TypeScript, JavaScript, Tailwind CSS, Material UI, Webpack, Vite, REST APIs, GraphQL
- **Backend:** Java, Go, .NET, .NET Core, Node.js, Python, Spring Boot, Flask, Django, FastAPI, Express.js, Microservices, gRPC, Message Queues, Apache Kafka
- **Databases:** PostgreSQL, MySQL, Aurora, DynamoDB, Redshift, Snowflake, SQL Server, MongoDB, Redis, Elasticsearch, OpenSearch, Cassandra
- **Data & ML:** Pandas, NumPy, scikit-learn, TensorFlow, PyTorch, SageMaker, Hadoop, Spark, EMR, Kinesis, NLP, Recommendation Systems, Deep Learning, Model Training & Deployment
- **AI & Automation:** OpenAI API, LangChain, LangGraph, Agentic AI Frameworks, RAG, Vector Databases, Pinecone, Embeddings, Prompt Engineering, Voice AI, GenAI, Personalization, Forecasting
- **Cloud & DevOps:** AWS, OCI, GCP, Terraform, CDK, Docker, Kubernetes, Jenkins, GitHub Actions, GitLab CI/CD, CircleCI, Ansible, NGINX, Observability

Professional Experience

Staff AI Full Stack Engineer / Tech Lead

Preferred Networks, Inc. | Apr 2021 – Present

- Built an enterprise RAG platform processing 10M+ documents with LangChain + React conversational UI
- Architected real-time inference serving on Kubernetes — 50K req/s, sub-200ms P99 latency
- Created internal model evaluation framework, cutting LLM deployment validation from days to hours
- Led cross-functional team of 6 ML engineers and 4 frontend engineers.

Senior AI Engineer

LINE Corporation | Jan 2018 – Mar 2021

- Shipped smart reply feature generating contextual suggestions for 200M+ users
- Built ML feature store and real-time prediction pipeline (Kafka → Spark → Redis)
- Developed annotation platform UI used by 50+ internal labelers, 3x throughput improvement

ML / Full Stack Engineer

Recruit Holdings Co., Ltd. | Apr 2015 – Dec 2017

- Built NLP-powered job matching algorithms for Indeed Japan, +25% match rate

- Developed recruiter dashboard (Vue.js + FastAPI) visualizing ML predictions
- Implemented automated model retraining pipelines with Airflow and MLflow

Software Engineer

Rakuten Group, Inc. | Apr 2013 – Mar 2015

- Developed product recommendation APIs using collaborative filtering, 100K+ RPM
- Built real-time search autocomplete with Elasticsearch and behavioral signals
- Migrated legacy product pages to React SPA with server-side rendering

Education

M.Eng. in Computer Science

The University of Tokyo — NLP Lab, Apr 2011 – Mar 2013

B.Eng. in Informatics

Kyoto University, Apr 2007 – Mar 2011