

Hiro Tanaka

AI Engineer (Senior)

Contact

Email: hiro.tanaka@example.com

Phone: +81 70-1234-5678

Address: Tokyo, Japan

<https://linkedin.com/in/hirotanaka> |

<https://github.com/htanaka> |

<https://hiroai.dev>

Date of Birth: 1988-02-04

Nationality: Japanese

Profile

Senior AI Engineer specializing in LLM systems, evaluation, and optimization for latency and cost at scale.

Skills

- PyTorch
- Distributed Training
- Optimization
- RAG
- Prompt Engineering
- Evaluation
- MLOps

Experience

Senior AI Engineer — NeuraWave (Tokyo, Japan)

Sep 2021 – Present

- Designed multi-tenant inference with KV-cache sharing.
- Built eval harness with rubric and human-in-the-loop review.

Tech: PyTorch, vLLM, Ray, CUDA, Weights & Biases

Projects

EvalLab — Python, Ray, Weights & Biases

Automated LLM eval pipeline with bias and safety checks.

<https://github.com/htanaka/evallab>

Feb 2022 – Present

Education

Master — University of Tokyo

Faculty: Computer Science • Major: Artificial Intelligence • CGPA: 3.8

2012 – 2014

Certificates

- AWS Machine Learning – Specialty

Awards

Innovation Award — NeuraWave, 2023

Latency reduction project

Languages

- Japanese — Native
- English — Fluent