

杨际舟

应用 AI / AI 产品工程师 | 求职方向: Applied AI · AI Product · FDE

日本东京 | yjz.haruka@gmail.com | www.harukayang.com | github.com/YANG-Haruka | linkedin.com/in/yang-haruka

个人概要

应用 AI 工程师兼项目负责人, 擅长把研究原型推进为可靠运行的系统。在 Corpy 主导某日本车企的端到端自动驾驶系统, 负责整体架构、模型与数据集成、Python 到 C++ 的跨语言传输、推理格式转换和持续交付; 业余独立构建真实业务使用的 SaaS、OCR、RAG 与语音 AI 产品。求职方向为应用 AI、AI 产品工程与 FDE。在留资格: 高度専門職1号, 永住申请中。

专业技能

AI 原生开发 Agent 编码 (Claude Code、Codex)、快速技术调研、生图/生视频、AI 幻灯片渲染成演示视频——短时间调研 SOTA 并落地为面向客户的完整 demo

语言 Python、TypeScript、C++、SQL、Dart

AI / ML / 机器人 PyTorch、端到端自动驾驶、计算机视觉 (检测/分割/跟踪/人物 Re-ID)、模型部署 (ONNX / TensorRT)、RAG、LLM Agent、模型评测、OCR / 语音、ROS 2、OpenGL、CUDA

LLMOps / MLOps 多供应商模型路由与配置化兜底链、自适应限流、结构化输出经代码侧校验、稠密+BM25 混合检索与重排、对人工基准集做模型/提示词 A-B 评测

后端 FastAPI、PostgreSQL (多租户行级安全)、SQLAlchemy / Alembic、SQLite、Node.js、OpenAPI 契约驱动

前端/移动 React、TypeScript、Vite、Flutter

云 / 运维 AWS (Lambda、DynamoDB、S3、CloudFront、CDK)、GCP (Vertex AI / Cloud Run)、Docker、Fly.io、CI/CD (GitHub Actions / GitLab CI)、Linux

工作经历

Corpy & Co., Inc. — 应用 AI 工程师 → 项目负责人

东京 | 2024.05 – 至今

某日本车企端到端自动驾驶系统——研究原型到实车运行

- 主导整体系统架构, 将系统从研究原型推进到实车实时运行。
- 参与部分模型训练, 负责数据前处理、后处理及最终推理格式转换; 参与推理优化, 关键指标持平前提下推理提速约 70%。
- 解决 Python 到 C++ 的跨语言数据传输问题, 引入 FlatBuffers 优化传输效率和运行性能。
- 晋升项目负责人 (团队 4 → 7 人), 负责需求、每周交付与模块协调——模块负责制使交付效率提升约 50%; 并参与 Cross-team 鱼眼人物 Re-ID 前期 R&D (人物分割 vs B-box 检测)。

AMIT株式会社 — 医疗 AI / 机器人工程师 (正式员工)

2022.04 – 2024.04

术前 CT 与术中超声融合的手术导航 / 机械臂辅助消融系统 —— 与早稻田大学联合研究项目

- 独立负责机械臂运动学与多空间坐标变换 (图像↔患者↔臂基座); ROS 2 自动跟踪控制; 肝脏模体与临床试验验证。
- 搭建医生在环的半自动标注流程 (约 8 千张), 训练医学图像分割模型至 IoU 约 0.89, 大幅降低人工标注成本。
- 使用 ONNX / TensorRT 完成模型转换与推理加速——受限硬件 (RTX 2060) 上提速约 70%, 同时保持识别效果。

代表项目

- EXA Restaurant** —— 多租户餐厅运营 SaaS, 独立开发, 川崎一家真实餐厅每天在用 (11 名员工), 部署在 AWS 上。「A 店看不到 B 店的数据」这条规则交给数据库自己执行, 而不是靠每次写查询都记得加过滤条件——65 张表、256 条行级安全策略, 每次 CI 从代码重建并在真实 Postgres 上重新核对。877 个自动化测试覆盖后端 / Web / 手机端; 发布前会从零重建一个全新实例, 让它跑通自己的全部检查才放行。另做了加密 NFC 打卡, 标签无法复制、刷卡记录无法重放 (NTAG 424 DNA)。
- LinguaHaru / HaruYukaFlow** —— 生产级文档 AI: 16+ 格式翻译、OCR/漫画/音频模式 (GitHub 270+ stars), 以及服务真实跨境业务的三模型小票 OCR 与库存/报关流程。
- 个人网站 (harukayang.com) + HaruAccompany** —— 自己做的作品集网站与其上的 AI 聊天助手, 部署在 AWS 且不需要常驻服务器: 静态页走 CDN, 请求由 Lambda 按次处理, 数据存 DynamoDB, 基础设施全部用代码 (CDK) 描述。后端零第三方依赖, 仅用 Node.js 自带能力写成。聊天助手会先检索我的经历与项目再作答, 回复边生成边逐字显示。

黑客松

- Gemini Gym** —— Google Cloud Gemini AI Agent 黑客松 (东京 2026.06, 团队): 负责后端与展示; 引擎 JS→Python 零差异移植、Google ADK 接入 Gemini。
- FASHINI** —— AI 虚拟试衣智能镜, DevOps × AI Agent 黑客松 2026 (StyleAI 团队, 评审中): 负责后端开发; 已上线 Cloud Run。

教育背景

电气通信大学 (东京) · 机械智能系统学专攻 · 硕士

2021.04 – 2023.03

电子科技大学 (成都) · 机械设计制造及其自动化 · 本科

2015.09 – 2019.06

其他

- 语言: 中文 (母语)、英语 (工作语言)、日语 (日常会话, 向商务水平进阶中)。