

FARROS ALFERRO

farros240802@gmail.com | linkedin.com/in/farrosalferro | github.com/farrosalferro

EDUCATION

Tohoku University

M.Sc. in Information Sciences (GPA: 3.8/4.0)

Sendai, Japan

Oct. 2024 – Sep. 2026

Sato-Yo Scholarship (2024–2026)

Tohoku University

B.Eng. in Mechanical & Aerospace Engineering (GPA: 3.7/4.0)

Sendai, Japan

Oct. 2020 – Sep. 2024

MEXT SGU Scholarship (2023–2024)

SKILLS

Programming: Python, SQL, C; Familiar: C++, Java, TypeScript, JavaScript, HTML

Backend & Data: FastAPI, Pydantic, Streamlit, MySQL, Qdrant, BeautifulSoup

Machine Learning: PyTorch, HuggingFace Transformers, OpenCV

Tools & Platforms: Git, Docker, Bash, SLURM, GCP, LangGraph, Unreal Engine, Blender

Spoken Languages: English (TOEIC 990), Japanese (JLPT N2), Indonesian (Native)

EXPERIENCE

Mercari, Inc.

Machine Learning Engineer Intern

Remote, Japan

Apr. 2026 – Jun. 2026

- **Multimodal Anomaly Detection:** Built a compact ~45M-parameter self-supervised image-text encoder (37% smaller, 2.35× faster than the production model), distilled from frozen foundation models, to flag prohibited marketplace listings at <0.1% positive rate.
- At the production alert budget (~1,000 reviews/day), matched the deployed baseline on the established category and improved precision/recall on a newly onboarded category from 42%/74% to 47%/84%.

Okatani - Computer Vision Lab., Tohoku University

Research Assistant

Sendai, Japan

Oct. 2024 – Mar. 2026

- **Autonomous Driving:** Built a custom CARLA asset integration pipeline to import 18 fully functional OOD obstacles with Unreal Engine and Blender; used the resulting benchmark to evaluate 17 SOTA end-to-end driving models across 24,225 episodes, showing that new obstacle shapes and sizes hurt driving performance ~2.7× more than texture or color changes.
- **Visual Search (MLLM):** Built and optimized a multimodal visual search pipeline combining open-vocabulary detection and prompting, reducing V* Bench processing time from 1472s to 329s at comparable accuracy (4.5× speedup); curated a 25K-image small-object benchmark with semi-automated Gemini/LLMDet annotation, yielding ~1.3M detections across 40K+ unique labels.
- **Reviewer** for CVPR '25, ICCV '25, and BMVC '26.

Kisuitech

Machine Learning Engineer Intern

Remote, Japan

Jul. 2024 – Aug. 2024

- Standardized apple detection datasets for training pipelines and built Python scripts using AWS S3 (boto3) to manage storage and ingestion for iterative model development.
- Accelerated image preprocessing on Jetson Nano by porting the pipeline from CPU to GPU with CuPy, reducing preprocessing time by 91%.

PUBLICATIONS

360° Image Perception with MLLMs: A Comprehensive Benchmark and a Training-Free Method

ECCV 2026

- H. Tran, V. Nguyen, **F. Alferro**, K. Liu, T. Okatani. Introduces 360Bench, a high-resolution 360° VQA benchmark, and Free360, a training-free scene-graph method. [arXiv:2603.16179](https://arxiv.org/abs/2603.16179).

PROJECTS

Fashion Recommender Agent | Pydantic, LangGraph, Qdrant, FastAPI, Next.js, Docker Dec. 2025 – Jan. 2026

- Built a chat-based fashion recommender agent with a FastAPI /chat endpoint and a Next.js UI; deployed locally with Docker (FastAPI + Qdrant).
- Implemented a retrieval workflow over ~100 personal wardrobe items stored in Qdrant, enabling recommendations and similarity search.
- Integrated tool-using behaviors (recommendation, vector search, virtual try-on, web search) and monitored agent traces with LangSmith.

HuggingFace Contributor | PyTorch, HuggingFace Transformers, OpenCV, GitHub Dec. 2023 & Nov. 2024

- Shipped new features to HuggingFace: integrated [NVIDIA Nemotron GGUF loading support](#) and contributed to the Fast Image Processor for Chameleon.
- Contributed improvements to the Computer Vision Community Course content covering [modern vision model](#) architectures.

Gecko (Bachelor Thesis) | PyTorch, HuggingFace Transformers, Gradio Jun. 2024 – Aug. 2024

- Developed Gecko, a training-free plug-and-play method to improve small-object visual search in high-resolution images for vision-language models.
- Curated Multi V* Bench (4 tasks × 20 samples; average resolution 1920×2475) and evaluated open-source and proprietary models.
- Released a public Gradio demo, model, and benchmark (website + code).