

Presentation Letter

I am an AI/ML Engineer and researcher with deep expertise in agentic AI systems, optimization, and MLOps, focused on compressing the cycle from research prototype to production deployment. I have architected production-grade multi-agent workflows using Google ADK and LangGraph, built retrieval-augmented generation systems, and deployed full-stack AI services with automated LLM-as-Judge evaluation harnesses. Throughout these projects, I led the architectural design, established foundational engineering standards to empower cross-functional teams, and guided engineering execution.

My academic research spans Bayesian optimization, Monte Carlo Tree Search, and multi-objective evolutionary algorithms, having produced internationally recognized results including a Best Student Paper Award at EMO 2019 and multiple peer-reviewed publications. My industry experience covers the complete research-to-production cycle. This includes designing scalable MLOps pipelines, conducting LLM fine-tuning research, and developing deep generative models at WEG where I mentored peer engineers and aligned stakeholders, as well as deploying low-latency voice AI and knowledge-grounded systems at Zazmic while coordinating lean development teams.

I hold a Master's in Engineering from Shinshu University as a MEXT Government Scholar and a Bachelor's in Electronic Engineering, with international academic experience across Japan, the United Kingdom, and Canada. I am driven to bring this unique blend of rigorous academic research and production engineering to a collaborative environment. My goal is to lead the development of next-generation intelligent workflows, bridge the gap between discovery and deployment, and help scale the collective impact of engineering and research teams.

Professional Experience

Zazmic Inc

United States of America (Remote)

AI/ML Engineer

Dec. 2025 – Present

- **Agentic AI Systems:** Architected production-grade multi-agent workflows using Google ADK and LangGraph, including ReAct agents for NL2SQL, Human-in-the-Loop (HITL) state management, and autonomous generators capable of ingesting diverse, multimodal unstructured data.
- **Real-Time Voice AI:** Engineered a low-latency, bidirectional voice agent utilizing WebSockets, FastAPI, and the Gemini Live API; implemented asynchronous tool definitions to support seamless, hands-free user interactions and real-time agent capabilities.
- **Retrieval-Augmented Generation:** Architected production RAG systems on GCP Vertex AI, grounding enterprise knowledge retrieval in live document corpora with citation-traceable outputs.

- **Evaluation & Agentic CI/CD:** Implemented LLM-as-Judge automated evaluation harnesses alongside dual-tracing (Langsmith/Langfuse + OpenTelemetry); deployed containerized agentic services via GitHub Actions CI/CD to GCP Cloud Run, compressing prototype-to-production cycles through automated quality gates.

• Dry Ground AI

United States of America (Remote)

AI Engineer

Mar. 2025 – Nov. 2025

- **AI Voice Agent Development:** Developed and deployed AI voice agents to answer real estate-related calls; integrated tools to retrieve up-to-date information, send messages and emails, and automate client interactions.
- **Client Profiling & Lead Generation:** Developed a data science workflow for classifying user profiles and identifying potential clients, supporting strategic sales initiatives.

• WEG

Jaragua do Sul, Brazil

Data Scientist

Jun. 2024 – Mar. 2025

- **Wind Turbine Anomaly Detection:** Deployed a production anomaly detection system on AWS SageMaker anticipating failures up to two weeks ahead and flagging ~70% of maintenance interventions; companion explainability API surfaced root-cause diagnostics for maintenance engineers.
- **End-to-End MLOps:** Built a reproducible MLOps pipeline (MLflow, Docker, Kubernetes, AWS) spanning automated training, versioning, deployment, and live monitoring via Grafana dashboards backed by InfluxDB, enabling iterative model improvements at production scale with full experiment lineage tracking.
- **Deep Generative Modeling:** Initiated PyTorch VAE research for wind turbine fault classification using latent-space health indicators and reconstruction-error thresholding over SCADA signals; architecture designed and early training conducted before project handoff.
- **LLM Fine-Tuning Research:** Led technical scoping to domain-adapt an open-weights LLM via Supervised Fine-Tuning (SFT) for NL-to-structured IoT workflow compilation; designed training corpus pipeline and execution-based evaluation harness; proposal approved by corporate leadership.

• Macnica DHW

Florianópolis, Brazil

Data Scientist

Nov. 2022 – Oct. 2023

- **Smart Sensors:** Developed an SVM-based classification pipeline for vibration sensor data; applied multi-objective Genetic Algorithms (NSGA-II) to optimize the trade-off between model accuracy and computational efficiency for edge hardware deployment.
- **Computer Vision APIs:** Built a dual-layer video anonymization API using OpenCV and DBSCAN clustering for motion-based person detection, with a pre-trained ResNet-50-FPN model (PyTorch) as a verification layer; optimized specifically for low frame-rate surveillance footage.
- **IoT Development:** Led end-to-end IoT product development with full Docker containerization of the edge-to-cloud stack (MQTT Broker, PostgreSQL, InfluxDB, Grafana, Python services); implemented edge computing including FFT calculations directly on Senspider hardware.

• SSE Gridtech

Curitiba, Brazil

R&D Engineer

Jan. 2016 – Aug. 2016

Internship

Oct. 2014 – Dec. 2015

- **Smart Grid & IoT Solutions:** Developed a long-range LoRaWAN-based communication system to optimize smart metering infrastructure, reducing GSM dependency.

Research Experience

- **University of Warwick, Warwick Business School** Coventry, United Kingdom
Research *Oct. 2019 – Apr. 2022*
 - **Field of Research:** Ranking & Selection; Bayesian Optimization; Reinforcement Learning.
 - **Research:** Investigated sample-efficient policy learning for automating constrained supply chain scheduling of biopharmaceutical products (personalized therapies); developed Multi-Objective MCTS (MO-MCTS) using marginal hypervolume as the selection criterion and integrated Evolutionary Algorithms to evolve base policies, combining Bayesian optimization with ranking-and-selection theory.
 - **Collaboration:** Under supervision of Prof. Juergen Branke, we collaborated with Prof. Chun-Hung Chen on two-stage Ranking and Selection (R&S) problems, which contributed to research extending Optimal Computing Budget Allocation (OCBA) and Expected Value of Information (EVI) algorithms to efficiently allocate finite stochastic simulation budgets on Bayesian approaches for optimal decision-making.
- **Shinshu University** Nagano, Japan
MSc Research *Apr. 2017 – Mar. 2019*
 - **Novel Method:** Developed a predictive model to approximate complex decision space geometries caused by difficult Pareto set topologies. The integration of this modeling as a mechanism into MOEAs proved to significantly improve solution distribution and convergence.
 - **Algorithm Analysis:** Analyzed Pareto dominance-based approaches on highly complex, multi-modal problems. Demonstrated that replacing standard operators with **Differential Evolution (DE) crossover** and neighborhood tournament selection substantially improves search ability and Pareto Front coverage.
 - **Collaboration & Recognition:** Collaborated with researchers across the **International Associated Laboratory LIA-MODO** (Shinshu University, Inria France, and UAM Mexico). Joint research on spline interpolation for bi-objective problems culminated in multiple publications and the **Best Student Paper Award at EMO-2019** (Michigan State University, US).

Education

- **Shinshu University** Nagano, Japan
Master of Engineering in Electronics and Information System *Apr. 2017 – Mar. 2019*
 - MEXT Monbukagakusho Scholarship (Japanese Government)
- **Concordia University** Montreal, Canada
Exchange Student, Electrical Engineering *Aug. 2012 – Aug. 2013*
 - Science Without Borders Scholarship (Brazilian Government)
- **Federal University of Technology - Parana** Curitiba, Brazil
Bachelor of Engineering in Electronics *Jan. 2010 – Jul. 2016*

Technical Expertise

Programming Languages:	Python, C, C++, Bash (Unix Shell).
ML & Deep Learning:	PyTorch, Hugging Face Transformers, scikit-learn, XGBoost.
LLM & Agentic Frameworks:	LangChain, LangGraph, Google ADK, Gemini Live API, OpenAI API, Anthropic Claude.
LLM Training & Alignment:	SFT, QLoRA, Unsloth.
Observability & Evaluation:	Langfuse, LangSmith, OpenTelemetry, Grafana.
MLOps & Deployment:	Docker, Kubernetes, MLflow, Git, CI/CD, FastAPI.
Embedded Systems:	ARM Cortex-M, Jetson Nano, Arduino, FreeRTOS, MQTT.
Data Engineering & Pipelines:	PostgreSQL, Redshift, InfluxDB.
Cloud Platforms:	AWS (SageMaker, ECR, Lambda), GCP (Cloud Run, Vertex AI).
Optimization:	Bayesian Optimization, Multi-objective Evolutionary Algorithms.
Operating Systems:	Linux (Fedora, Ubuntu), Windows.

Main Publications

- **Y. Marca**, H. Aguirre, S. Zapotecas, A. Liefoghe, B. Derbel, S. Verel, and K. Tanaka. Approximating Pareto set topology by cubic interpolation on bi-objective problems. 10th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2019), Lecture Notes in Computer Science (LNCS), Michigan, USA. (**Best Student Paper Award**)
- **Y. Marca**, H. Aguirre, S. Zapotecas-Martinez, A. Liefoghe, B. Derbel, S. Verel, and K. Tanaka. MOEA with Cubic Interpolation on Bi-objective Problems with Difficult Pareto Set Topology (2019). doi:10.11394/tjpnsec.10.12.
- **Y. Marca**, H. Aguirre, S. Zapotecas, A. Liefoghe, B. Derbel, S. Verel, and K. Tanaka. NSGA-II with Spline Interpolation on Bi-objective Problems with Difficult Pareto Set Topology. JPNSEC 2018 Symposium on Evolutionary Computation, Fukuoka, 2018. (**Young Research Award**)
- **Y. Marca**, H. Aguirre, S. Zapotecas, A. Liefoghe, B. Derbel, S. Verel, and K. Tanaka. MOEAs on Problems with Difficult Pareto Set Topologies. IEICE Shin-etsu Branch IEEE Session, Niigata University, 2018, p. 169.(**Young Research Award**)
- **Y. Marca**, H. Aguirre, S. Zapotecas, A. Liefoghe, B. Derbel, S. Verel, and K. Tanaka. Pareto dominance-based MOEAs on problems with difficult pareto set topologies. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '18). ACM, New York, NY, USA, 189-190.

Languages

Portuguese:	Native
English:	Fluent
Japanese:	Conversational