Hotsuyuki Kawanishi

hotsuyuki.kawanishi@gmail.com | +1 (786) 674-6922 | linkedin.com/in/hotsuyuki | github.com/hotsuyuki

SUMMARY

Software Engineer with 4 years of experience in Autonomous Robot (Motion Planning, Perception, Localization & Mapping) and Machine Learning (Large Language Model, Computer Vision). Set to graduate in Dec 2025, eligible for internship in Summer 2025.

EDUCATION

University of Miami, Coral Gables, Florida | M.S. in **Software Engineering** (GPA: 4.00) **Dec 2025**
Relevant Courses: Machine Learning, Neural Networks, Algorithm Design and Analysis, Software Architecture, Software DevOps

Kanazawa University, Kanazawa, Japan | M.Eng. in **Robotics** (GPA: 3.93) **Mar 2020**
Relevant Courses: Intelligent Robotics, Dynamics and Control, Master's Thesis (on Localization & Mapping for Self-Driving Car)

Chiba University, Chiba, Japan | B.Eng. in **Mechanical Engineering** (GPA: 3.87) **Mar 2017**
Relevant Courses: Linear Algebra, Calculus, Numerical Computation, Analytical Mechanics, Control Theory, Mechatronics

National Institute of Technology Numazu College, Numazu, Japan | A.Eng. in **Computer Science** (GPA: 3.81) **Mar 2015**
Relevant Courses: Data Structures and Algorithms, Programming, Operating Systems, Computer Graphics, Computer Simulation

EXPERIENCE

University of Tokyo AI Research Institute (Matsuo Lab), Tokyo, Japan **Jan 2024 – Aug 2024**
Machine Learning Engineer, LLM (part-time)

- Rapidly acquired necessary expertise in LLM within 1 month despite working full-time in a different field (Self-Driving Car), and built an LLM distributed training pipeline with Microsoft Megatron-DeepSpeed and Hugging Face Transformers, achieving successful training of an OpenAI GPT-based 10B generative model on 160 Nvidia H100 GPUs via Google Cloud GPU cluster.
- Led a lecture on pre-training in an online LLM course with 4,000+ participants, accomplishing the highest Net Promoter Score.

Woven by Toyota, Tokyo, Japan **May 2021 – Aug 2024**
Software Engineer, Self-Driving Car - Motion Planning (full-time)

- Enhanced safety validation for Level 4 driverless autonomous vehicles by designing and implementing a prediction evaluator with Agile methods and software development life cycle best practices, e.g. object-oriented design, code reviews, and testing.
- Decreased driver interventions in JP by 30% by debugging the optimization-based trajectory planning modules with teams in the US, e.g. triaging log data, identifying root causes on large codebases, articulating solutions, and managing car experiments.
- Streamlined and accelerated simulation scenario generation process for the motion planning team by 3x efficiency by creating automation scripts that leverage Applied Intuition API and by leading cross-functional collaboration with teams in the US.

Kanazawa University Advanced Mobility Research Institute, Kanazawa, Japan **Apr 2020 – Mar 2021**
Research Scientist, Self-Driving Car - Perception (full-time)

- Improved map data collection process for autonomous vehicles by 2x efficiency by creating a real-time map generation tool.
- Developed perception features by applying research theories to practical applications in a fast-paced environment, including a speed bump detector using LiDAR point cloud and statistical approaches (successfully delivered to a customer company) and a lane line visibility checker using LiDAR-camera sensor fusion and linear algebra (published in an international journal paper).
- Mentored 4 lab students, resulting in 1 paper publication at IEEE IV and 1 patent filing, by providing hands-on guidance.

Cultural Vistas, New York City, New York **Jul 2017 – Dec 2017**
Software Engineer, Web (internship)

- Developed and maintained the web front-end and back-end in a diverse, multicultural environment of 5 software engineers.

SKILLS

Languages / Technologies: C, C++, C#, Python, Shell Script, JavaScript, TypeScript, PHP, SQL, PyTorch, TensorFlow, DeepSpeed, Transformers, W&B, Slurm, Jupyter Notebook, Unity, React, FastAPI, AWS, Azure, GCP, Docker, Jenkins CI, Bazel, Git, Linux, CLI

PROJECTS

LLM Distributed Training Pipeline | *Python, Shell Script, DeepSpeed, Transformers, Distributed Systems* **Jan 2024 – Apr 2024**

- Scalable distributed training pipeline for LLMs on GPU clusters using DeepSpeed ZeRO, 3D Parallelism, and Transformers.

HuggingGPT using GPT Function Calling | *TypeScript, React, Python, FastAPI, Azure, OpenAI API* **Jul 2023 – Jul 2023**

- Agentic AI prototype to interact with various Hugging Face APIs through natural language utilizing GPT Function Calling.

C++ Deep Learning Framework from Scratch | *C++, Bazel, Object-Oriented Programming, Unit Testing* **Jun 2022 – Aug 2022**

- PyTorch-like deep learning framework architected and implemented from scratch in C++, e.g. layers, optimizers, and autograd.

Object Detection Model in PyTorch C++ with TensorRT | *C++, PyTorch, TensorRT, Conv Neural Net (CNN)* **Dec 2020 – Jan 2021**

- SSD300 in PyTorch C++ (instead of Python) with TensorRT for faster performance and lower inference latency in production.

Robotics Algorithms in Jupyter Notebook | *Python, Robotics, SLAM, Localization, Path Planning* **Aug 2020 – Sep 2020**

- Python codes for robotics algorithms, e.g. Simultaneous Localization & Mapping (SLAM), localization, and path planning.