

Hotsuyuki Kawanishi

hotsuyuki.kawanishi@gmail.com | +1 (786) 674-6922 | [linkedin.com/in/hotsuyuki](https://www.linkedin.com/in/hotsuyuki) | github.com/hotsuyuki | hotsuyuki.github.io

SUMMARY

Software Engineer with 4+ years of professional experience in Machine Learning (Large Language Model, Computer Vision) and Self-Driving Car (Localization & Mapping, Perception, Motion Planning). Eligible for US internship under CPT in Summer 2025.

EDUCATION

University of Miami, Coral Gables, Florida **Dec 2025**
Master of Science in **Software Engineering** **GPA: 4.00**
Relevant Courses: Machine Learning, Statistical Learning, Neural Networks, Software Architecture, Software DevOps

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

Chiba University, Chiba, Japan **Mar 2017**
Bachelor of Engineering in **Mechanical Engineering** **GPA: 3.87**
Relevant Courses: Linear Algebra, Calculus, Numerical Computation, Analytical Mechanics, Control Theory, Mechatronics

National Institute of Technology Numazu College, Numazu, Japan **Mar 2015**
Associate of Engineering in **Computer Science and Engineering** **GPA: 3.81**
Relevant Courses: Data Structures and Algorithms, Programming, Operating Systems, Computer Graphics, Computer Simulation

EXPERIENCE

University of Tokyo Matsuo Lab, Tokyo, Japan **Jan 2024 – Aug 2024**
Machine Learning Engineer, LLM (part-time) | Python, Shell Script, Pre-training, Fine-tuning

- Built an LLM distributed training [pipeline](#) using 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 (full-time) | C++, Python, Prediction, Motion Planning

- 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 fixing bugs in decision making and trajectory planning with teams in the US, e.g. triaging failures, identifying root causes on a large codebase, articulating solutions, and coordinating vehicle experiments.
- Streamlined and accelerated simulation scenario generation process for the motion planning team by 3x efficiency by creating automation scripts leveraging Applied Intuition API and collaborating with multiple cross-functional teams in the US.

Kanazawa University Autonomy Lab, Kanazawa, Japan **Apr 2020 – Mar 2021**
Research Scientist, Self-Driving Car (full-time) | C++, Python, Localization & Mapping, Perception

- 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 findings into real-world practices in a fast-paced R&D environment:
 - Speed bump detection with LiDAR point cloud and probabilistic method, delivered to the customer company successfully.
 - Lane line visibility check with LiDAR-camera sensor fusion and linear algebra, published in an international journal [paper](#).
- Successfully mentored 4 lab students, resulting in 1 [patent](#) filing and 1 [paper](#) publication, by providing hands-on guidance.

Cultural Vistas, New York City, New York **Jul 2017 – Dec 2017**
Software Engineer, Web (internship) | JavaScript, PHP, SQL, Full-stack

- Developed and debugged the web front-end and back-end in a diverse, multicultural environment of 5 software engineers.
- Granted a \$10,000 Japanese government scholarship for demonstrated leadership and commitment to learning tech in the US.

SKILLS

Programming Languages: C++, C, C#, Python, Shell Script, JavaScript, TypeScript, PHP, SQL, HTML, CSS

Frameworks / Technologies: PyTorch, TensorFlow, Keras, DeepSpeed, Hugging Face Hub, Transformers, W&B, CUDA, ROS, Jupyter Notebook, NumPy, Matplotlib, Unity, Niantic ARDK, React, FastAPI, AWS, Azure, GCP, Docker, Jenkins, Bazel, Git, Linux

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 by GPT Function Calling](#) | TypeScript, React, Python, FastAPI, Azure, OpenAI API **Jul 2023 – Jul 2023**

- Agentic AI web app to interact with various Hugging Face APIs through natural language by 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.