


# Bin-Lun Li

 [mike911209](#) |  [my site](#) |  [mike911209@gmail.com](mailto:mike911209@gmail.com) |  +886 982-880-498

## EDUCATION

---

**National Tsing Hua University (NTHU)**

Sep 2021 - Jun 2025

B.S. in Computer Science

Advisor: Professor **Jerry Chou**

Average **GPA 4.27/4.3** on last 2 semesters.

## EXPERIENCE

---

**Machine Learning Engineer Intern - Lasertec Taiwan**

Sep 2024 - Dec 2024

Pytorch/Triton/CUDA/Linux

- **Accelerate** morphological operation via parallel programming.
- **Build** industrial-oriented software on GPU cluster.

**Teaching Assistant of Operating System**

Sep 2024 - Dec 2024

C/C++/Operating System/Linux

- **Designed** the specifications for Operating System implementation.
- **Setup** the server environment for class students.

**Teaching Assistant of Hardware Design and Lab**

Sep 2024 - Dec 2024

Verilog/Hardware Design

- **Developed** the specifications for hardware design problem.
- **Resolved** issues caused by both hardware and software, thereby enhancing my debugging skills.

## AWARDS

---

**Second Prize - 2024 Meichu Hackathon**

[Link to Project](#)

Python/Flask/Backend/LLM

- **Won** 2nd prize among 230 people, first prize among Logitech group.
- **Multi-agent chatroom**, each with a distinct personality. You can gain insights from the interactions between your AI agents and engage with them.

## PROJECTS

---

**Scalable Serving System for LLM with Kubernetes and NVIDIA MIG**

Go/Kubernetes/Prometheus/Grafana/NVIDIA MIG/System Design

- **Utilized** NVIDIA MIG to allocate GPU resources, effectively addressing resource fragmentation and resource contention issues.
- **Dynamically adjusted** resource allocation based on runtime resource usage of LLM inference server, leveraging NVIDIA MIG to ensure optimal resource utilization.
- **Monitored** resource usage in real-time and visualized data through Grafana dashboards.

## SKILLS

---

**Programming Languages** C/C++, Python, Java, Javascript, Verilog, Go

**Tools & Libraries** Git, Unix-like shells, Docker, Kubernetes, CUDA, MPI, OpenMP

**Languages** TOEIC Listening and Reading Test 915/990