Caiwu Chen

(929) 605-8323 | cc4786@columbia.edu | **US Citizenship**
Linkedin: https://www.linkedin.com/in/caiwuchen/ | Github: https://github.com/CaiwuChen

EDUCATION

Columbia University New York, NY
MS in Computer Science, Software System track Expected Dec 2025
- Course: Cloud Computing, Operating Systems, Embedded Systems, UI Design, Distributed Systems, Advanced Software Engineering, Parallel Functional Programming, Programming Language and Translators

BA in Computer Science, Intelligence track Feb 2024
- Course: Advanced Programming, Artificial Intelligence, Databases, Natural Language Processing, Computational Robotics, Machine Learning, Analysis of Algorithms, Computer Vision, Neural Network and Deep Learning

LANGUAGES AND IT SKILLS

- Programming Languages: Python, Java, C/C++, SQL, HTML, JavaScript, Haskell, Golang, SystemVerilog
- Technology & Tools: Linux, Git, AWS, MeshLab, MongoDB, GCP, LaTeX, TensorFlow, TFLite, Spring Boot

WORK EXPERIENCE

Columbia Build Lab New York, NY
Distributed Data & Management Platform Jan 2025 - Aug 2025
- Architected and enhanced a scalable management service on compute instances in GCP, engineering it for high availability and fault tolerance by deploying stateful replicas across multiple Availability Zones.
- Optimized a managed SQL service to cut query latency by 40% for large-scale analytics and designed data layers for horizontal scalability.
- Boosted search relevance by 25% by building a serverless TF-IDF pipeline, performing dynamic query expansion on data in object storage.

LaGuardia Community College, CUNY New York, NY
Undergraduate Researcher Jan 2021 - June 2021
- Proposed a schematic sandbox model describing nucleon-nucleon correlations in wave functions.
- Developed C++ simulation of nucleon dynamics and analyzed separation energies for O-16 nuclei.
- Interpreted experimental data using Hamiltonian and Schrödinger equations.

RESEARCH & PROJECTS

Synchronic Automatic Sign Language Recognition Jan 2025 - May 2025
- Architected and developed a real-time ASL recognition system, processing live video streams via a dual-input CNN-MLP model and leveraging MediaPipe for high-fidelity hand landmark extraction.
- Engineered a temporal smoothing filter (sustained-confidence logic) to stabilize predictions across frames, reducing output jitter and cutting false positive recognitions by 20%.
- Developed and modified a cross-platform deployment pipeline for both desktop (TensorFlow) and mobile (TFLite), achieving an end-to-end inference latency of under 150ms on live camera feeds.

Full-Stack SaaS Scheduling Platform Sep 2024 - Dec 2024
- Led a 5-person team to architect, develop, and deploy an end-to-end SaaS scheduling platform, built with Java, Spring Boot, and JPA on a 3-tier architecture backed by a relational SQL database.
- Engineered the backend's core scheduling engine with RESTful APIs, featuring intelligent timeslot merging and conflict resolution to handle complex availability scenarios.
- Constructed client portals, implementing role-based access control to deliver distinct workflows and data user roles.
- Established a unified CI/CD pipeline using GitHub Actions for the entire system, automating builds, testing (JaCoCo for coverage), and deployment to a live cloud environment.

Comparative Analysis of Vision Architectures for Fine-Grained Classification Sep 2023 - Dec 2023
- Conducted a comparative analysis of ResNet50 and Vision Transformer (ViT) on a small, low-resolution dataset for a multi-label, fine-grained image classification task.
- Fine-tuned a pre-trained ResNet50 using advanced hyperparameter optimization and custom data augmentation, achieving significant validation loss reduction and setting a new performance benchmark.
- Demonstrated ResNet50 achieved over 15x higher accuracy on sub-class labels than ViT (4.1% vs 0.27%), a result attributed to significant ViT overfitting identified through loss curve analysis.