

BHANUJA KARUMURU

📍 New York, USA 📞 +1 3474459258 🌐 [linkedin.com/in/bhanujakarumuru](https://www.linkedin.com/in/bhanujakarumuru) 🐙 github.com/Bhanuu01 ✉ bk3170@nyu.edu

EDUCATION

New York University, MS in Computer Engineering New York, USA | Expected May 2027
Relevant Coursework: High Performance Machine Learning, MLOps, Data Center & Cloud Computing, Systems Design
NIT Sikkim, B.Tech in Electronics and Communication Engineering Sikkim, India | Dec 2020 – May 2024
Relevant Coursework: Data Structures & Algorithms, Machine Learning, Neural Networks, Deep Learning, Computer Networks

WORK EXPERIENCE

Incoming Software Development Intern - Amazon.com Inc. Seattle, USA | Starting June 2026
Software Development Engineer, SalesUp - Yellowlake Technologies Services Private Limited Kolkata, India | July 2024 – July 2025
(Promoted from Software Engineering Intern)

- Built Flask-based microservices for a CRM platform, improving throughput by 40% through async task queues (Redis/Celery), REST API refactoring, and PostgreSQL query optimization.
- Developed an LLM-powered email campaign feature using RAG and vector embeddings, increasing user engagement by 25%.
- Containerized applications with Docker and automated Jenkins deployments in AWS/Linux environments, reducing release overhead by 30%.
- Profiled and restructured async request handling, reducing P99 API latency by 25%.

RESEARCH EXPERIENCE

Student Researcher (TReX Fellow) - NYU Tandon School of Engineering New York, USA | Starting June 2026

- Developing ML/DL-based computer vision pipelines for object detection, feature extraction, and geotagging from high-resolution historic maps, enabling GIS-compatible spatial data integration for urban infrastructure analysis under Prof. Debra Laefer.

AI/ML Infrastructure & Computer Vision Researcher, NYU RoboMaster (Team Ultraviolet) New York, USA | Jan 2026 – Present

- Built and integrated real-time object detection and tracking pipelines for autonomous robots in NYU's VIP RoboMaster program, supporting vision model training and on-robot deployment with controls and hardware teams.

Graduate Research Contributor, Secure Systems Lab, New York University New York, USA | Sep 2025 – May 2026

- Contributed to virtualization-based OS security, implementing hypervisor-level privilege controls and kernel isolation to reduce cross-VM attack surfaces in multi-tenant environments. Integrated namespace confinement and resource sandboxing primitives.

PROJECTS

Distributed Log Aggregation System

- Built a distributed log aggregation system in Python with producer-consumer concurrency, B-tree indexing, and a Redis caching layer; REST APIs for log ingestion and filtering achieve sub-50ms query response under high-throughput writes.
- Deployed on AWS EC2 with Docker for horizontal scaling; validated end-to-end correctness with pytest unit and integration tests.

Microservice-Based Task Orchestration Platform

- Architected a distributed task platform with priority queuing, retry logic, and audit logging using Flask, Supabase, and async task queues; designed for horizontal scalability under sustained load.
- Deployed Dockerized services on AWS EC2 with GitLab CI/CD, enabling zero-downtime rollouts and horizontal scalability.

End-to-End Fraud Detection Platform

- Engineered a high-throughput ETL and processing pipeline with containerized microservices on AWS EC2, continuous deployment via Jenkins, and rollback support for versioned model updates.
- Integrated ensemble models (logistic regression, random forest, gradient boosting) into the serving layer, achieving AUC 0.96 and reducing false positives by 35% through feature engineering and threshold tuning.

Distributed Rate Limiter

- Implemented Token Bucket and Sliding Window Log algorithms in Python with thread-safe atomic operations and a Redis-backed shared state layer enforcing consistent limits across instances with sub-5ms overhead.
- Validated correctness and thread-safety with pytest unit and integration tests simulating burst traffic; added structured logging and metrics export for observability.

TECHNICAL SKILLS

Programming & Systems: Python, C/C++, Java, JavaScript, SQL, Bash, CUDA, Git, Linux
Systems & Design: Distributed Systems, Object-Oriented Design, Concurrency, Multithreading, Event-Driven Architecture, REST API Design
Backend & Frameworks: Django, Flask, FastAPI, PostgreSQL, MongoDB, Supabase, Microservices, Redis
Cloud & DevOps: AWS (EC2, S3, Lambda), Docker, Kubernetes, Jenkins, CI/CD
AI/ML & Data Science: PyTorch, TensorFlow, HuggingFace Transformers, Scikit-learn, Pandas, NumPy, pytest

PUBLICATIONS

Fusion of Data Augmentation for Improved Dysarthria Severity Classification, Journal of Signal Processing Systems, Springer, Accepted 2026
In-Domain Data Augmentation for Dysarthria Severity Classification, SPCOM 2024