

## Education

☎ (+1) 347-741-3100 | ✉ ax2135@nyu.edu | 📧 /Aku2611 | 📄 ankit-chahar2611 📍 Brooklyn, New York

## New York University

New York City, U.S.A

MASTER OF SCIENCE IN COMPUTER ENGINEERING

August 2024 - May 2026

- Merit Scholarship Recipient · Coursework: Machine Learning, Deep Learning, High-Performance Machine Learning, Neural Networks, Big Data

## R.V. College of Engineering

Bengaluru, India

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

August 2019 - May 2023

- Courses: Software Engineering, Operating Systems, Machine Learning; Advanced Networking, object-oriented programming (OOP)

## Skills

- **Languages:** C/C++, C, Java, Python, JavaScript, TypeScript, SQL, Bash, MATLAB, R · **IDEs:** Visual Studio, VS Code, Eclipse, IntelliJ IDEA, PyCharm
- **ML/AI:** PyTorch, TensorFlow, scikit-learn, NumPy, pandas, OpenCV, Hugging Face, MLflow, MLOps, ONNX, TorchScript, TFLite, Quantization, Pruning, Edge AI
- **Data/Big Data:** Apache Spark, Kafka, Hadoop, Airflow, Flink, ETL Pipelines, Streaming, Data Lakes, Data Warehousing, Snowflake, Databricks, Hive, Pig
- **Cloud/Backend:** RESTful APIs, Microservices, Node.js/Express, Spring Boot, FastAPI, GraphQL, PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch, Docker, Kubernetes, AWS (EC2, S3, EKS, Lambda, CloudWatch, DynamoDB), GCP, Terraform, CI/CD (GitHub Actions, Jenkins, ArgoCD), (Raspberry Pi), CUDA, OpenCL
- **Frontend:** React, Redux, React Native, Angular, HTML5, CSS3, TypeScript, Webpack/Vite, Jest, React Testing Library, Cypress, Storybook, Figma, Bootstrap
- **Tools/DevOps:** Linux, Git/GitHub, Grafana, Prometheus, Kibana, Splunk, JIRA, Confluence, Postman, Swagger, Tableau, Power BI, MATLAB Simulink

## Work Experience

## New York University

New York City, U.S.A

RESEARCH ASSOCIATE INTERN

May 2025 – Aug. 2025

- Engineered and deployed **MCUNet-in4** on Raspberry Pi 4B using **PyTorch, ONNX, and TensorFlow Lite**, achieving a **50%** model-size reduction (**11.67MB** → **5.56MB**) and **19.4% FLOPs reduction** while sustaining **91%+ Top-1 accuracy** on CIFAR-10.
- Implemented **block-aware soft pruning** (40% sparsity) and **iterative channel pruning** (56.6% removal), shrinking parameters from **1.41M** → **0.61M** with **< 4%** accuracy drop; benchmarked real-world performance 88ms/batch latency, 45–55°C temps, power usage), yielding a +2.2% runtime speedup and demonstrating practical **TinyML deployment** for IoT devices.

## MakeMyTrip Co., Ltd.

Bengaluru, India

SOFTWARE ENGINEER

Jan. 2023 – May. 2024

- Designed and developed web & mobile frontends with **React, Redux, React Native, HTML5/CSS3, JavaScript, and TypeScript**, delivering new features, UI revamps, and performance optimizations for large-scale production systems serving **50M+ monthly active users**.
- Delivered **10+ high-impact components** and maintained **20+ reusable, responsive React components** across **5+ codebases**, accelerating feature rollout and ensuring a consistent user experience across apps generating **billions in annual GMV**.
- Architected and deployed **microservices-based backend systems** using **Docker, Kubernetes, and AWS (EC2, S3, Lambda, CloudWatch)**, improving system reliability by **30%** and reducing infrastructure costs by **20%** through containerization, CI/CD, and cloud-native optimizations.

## Samsung Research Institute Inc.

Bengaluru, India

SOFTWARE ENGINEER INTERN

May 2021 – Jan. 2022

- Developed Android features leveraging the **MVP design pattern**, Jetpack architecture, and Google I/O 2017 best practices for scalable mobile applications.
- Delivered **Dynamic Delivery & Instant Apps**, reducing APK size by **60%**, accelerating app load times, and boosting installation rates in the App Store.
- Engineered a secure, high-performance **payments module** with multiple **Payment Gateways**, ensuring compliance, scalability, and seamless user experience.

## M2R Technomations

New Delhi, India

MACHINE LEARNING INTERN

May 2020 – Aug. 2020

- Applied **PySpark, NLP, and Machine Learning** to analyze large-scale customer-care conversations, automating insights and achieving significant **cost optimizations** in support operations.
- Implemented a high-performance **SPH fluid simulation** with in-house GPU pipeline and OpenGL compute shaders, rendering **200K+ particles at 30 FPS** in an Android app powered by Java's Filament rendering engine.

## Projects

## INSTANT APPLICATION FOR SAMSUNG STORE

GitHub

- Engineered an instant app for the Samsung Store; achieved **–25% APK size** and **–15% load time**, driving **+18% engagement**.
- Implemented Dynamic Delivery & Instant Features, performing R&D on Android App Bundles and Google Play Console optimization.

## JAILBREAKING DEEP MODELS

GitHub

- Designed adversarial attacks (FGSM, PGD, targeted patch) on **ResNet-34**, reducing Top-1 accuracy from 70.4% → 0.2% in 1 minute on 8k-image ImageNet subset.
- Built targeted patch attack pipeline (nearest-neighbour targets, optimal patch placement, momentum), forcing chosen misclassifications in **76.8%** of cases.
- Conducted cross-architecture transfer to **DenseNet-121**, revealing 59%+ Top-1 retention and architecture-dependent robustness gaps.

## AWS-DRIVEN CONTAINER ORCHESTRATION AND CI/CD IMPLEMENTATION

GitHub

- Designed and deployed a fully containerized app on **AWS EKS** with **Ingress, Load Balancing, HPA** for dynamic scalability and traffic management.
- Automated **CI/CD pipelines using GitHub Actions**, building/pushing Docker images to **Amazon ECR** and deploying to Kubernetes clusters.
- Implemented monitoring/logging with **Amazon CloudWatch**, improving performance insights and security env variable management and secure API routing.

## Research and Publications

- “Stroke Prediction System Using Machine Learning Algorithm.” In: **ICMCSI**, 2021. Published in Springer Lecture Notes.
- “Student Performance Prediction using Ensemble Technique.” In: **ICACCS**, 2022. Published in IEEE Xplore.
- “The Implementation of Big Data with Cloud and Edge Computing in Enhancing the Smart Grid Information Processes Through SEM Model.” In: **ICACCS**, 2022. Published in IEEE Xplore.