

Ritvik Rangaraju

ritvik@ritvik.it • (609) 921-4921 • ritvik.it • linkedin.com/in/ritvik-rangaraju

EDUCATION

University of Maryland - College Park

Bachelor of Science in Computer Science, Minor in Technology Innovation Leadership

College Park, MD

August 2022 – May 2026

Coursework: Multimodal Deep Learning, Cloud Computing, Computer & Network Security, Computer Vision, Database Design, Machine Learning, Human-Computer Interaction, Applied Cryptography, Advanced Data Structures, Compilers

Extracurriculars: *Trip Leader @ Adventure Program* (Hiking, Backcountry Camping, Backpacking, Caving, Kayaking), *Resident Assistant @ Department of Resident Life*, *Photographer @ Indian Student Association*

PROFESSIONAL EXPERIENCE

Amazon Web Services

May 2025 — August 2025

Software Development Engineering Intern

Seattle, WA

- Architected a zero-downtime update system in Java for a managed database service, implementing rolling updates to avoid cluster-wide outages, improving availability from 95% to 99.9% – directly impacting 60+ enterprises across 10,000+ clusters.
- Designed batched update architecture that parallelizes node updates, reducing downtime from 15 minutes to <30 seconds per node and decreasing system overhead by up to 35%.
- Reduced on-call burden by 20%, saving 150+ developer hours annually by preventing update-related production failures that previously caused 120+ customer tickets quarterly.

Peraton Labs

June 2024 — August 2024

Software Engineering Intern

Basking Ridge, NJ

- Developed an autonomous training pipeline for on-device deep learning models on tactical edge assets, improving location and network status prediction accuracy by 20% – enabling high-fidelity data and insights for real-time mission planning.
- Implemented a continuous re-training framework driven by live feature workflows – sensor, topographic, and meteorological data – updating deployed models in real time and reducing long-term predicted path deviance by 15%.
- Developed a compact, high-efficiency model serialization framework that reduced storage size by 40% and inference latency by 25%, enabling high-performance model serving on resource-constrained edge devices with minimal system footprint.

SEQ Technology

May 2023 — August 2023

Software Engineering Intern

New York, NY

- Engineered a natural-language analytics tool that converts English business questions into specialized SQL queries for client data warehouses using a database schema-aware RAG pipeline and LLM orchestration layer, reducing analysis time by 45%.
- Fine-tuned the FLAN-T5 LLM on healthcare/banking terms and entity extraction, improving conversion accuracy by 25% and raising one-shot task success to 85% via an active-learning evaluation loop from user edits.
- Delivered full stack: React frontend, Express gateway, FastAPI model serving microservice, and BigQuery execution.

PROJECTS

Des Rangila [Digital Passport System] | Next.js, Firebase, Twilio, InsightFace, AWS (EC2)

March 2026 — April 2026

- Solo-built and shipped a comprehensive live event management platform for a 600+ attendee cultural festival – owning product design, system architecture, event logistics, and live operations end-to-end.
- Designed the attendee experience – digital QR passports in Apple Wallet, Twilio SMS notifications, and station check-ins – deploying self-hosted, opt-in facial recognition via InsightFace to power event-wide identity flows with live model inference.
- Operated platform through live production from real-time admin dashboard: processing 5,000+ passport redemptions across 14 concurrent stations and diagnosing and fixing production incidents mid-event with near-zero downtime.

PassDrop | React, Express.js, WebCrypto, WebAuthn, AWS (App Runner, DynamoDB, CloudWatch)

February 2026

- Created a passwordless, end-to-end encrypted file sharing web app with client-side AES-GCM encryption, with keys generated entirely in the browser and kept in the URL fragment, only storing ciphertext and metadata in the backend.
- Implemented passwordless sign-in with WebAuthn passkeys for link owners and private link access for recipients, with features including timed link expiry, one-time downloads, and sender-controlled at-will revocation.
- Deployed a stateless API on AWS App Runner along with CloudWatch alarms for throughput, error rate, and storage usage.
- Added auditable, idempotent operations to prevent duplicate links and partial writes, tracking file metadata via DynamoDB.

TECHNICAL SKILLS

Languages: Java, Python, JavaScript, Go, Rust, C, Bash, SQL

Full-Stack: React, Next.js, Express.js, FastAPI, Node.js, Flask, Django, Firebase, OAuth

AI / ML Systems & Dev: TensorFlow, PyTorch, RAG, LLM fine-tuning, evaluation, Pandas, OpenCV, Claude Code, Codex, Cursor

Infrastructure & Cloud: AWS (EC2, Lambda, S3, DynamoDB, App Runner, CloudWatch), Docker, Kubernetes, PostgreSQL, Linux, Git