Amit Krishna Anantha Kumar

Town Hall Road, Koovappadam, Ernakulam, Kochi, Kerala 682002
+919746096525 | amit.ananthkumar@gmail.com | linkedin.com/in/akri16 | github.com/akri16 | leetcode.com/akri16

EDUCATION

Vellore Institute of Technology June 2019 - June 2023
Bachelor of Technology in Computer Science *CGPA: 8.37/10*

- **Relevant Coursework:** Data Structures and Algorithms, Operating Systems, Computer Networks, Object-Oriented Programming, Database Management Systems, Microprocessors and Microcontrollers, Internet of Things, Social Information Networks

PROFESSIONAL EXPERIENCE

Senior Software Engineer – ModemAI, Qualcomm, Hyderabad Nov 2026 – Present

Agentic Data Analysis

- Developed an end-to-end agentic AI system (ReAct, LangChain) for natural-language analysis of **petabyte-scale** modem data, leveraging AWS Glue/Athena for large-scale data discovery and querying, with automated analysis and QuickSight visualization
- Leveraged a custom dependency graph framework, Qwen embedding model, BM25, and LanceDB to enable **agentic graph RAG** across **100+ technical documents** and **10M+ lines** of modem code, supporting retrieval and reasoning across implementation details, APIs, specifications, and domain-specific knowledge.
- Deployed the system for engineering workflows used by **hundreds of engineers**, supporting **10K+ queries per week** and reducing engineers' Time-to-Insight (TTI) by **60%**, based on user feedback and evaluation.

Software Engineer – ModemAI, Qualcomm, Hyderabad Dec 2024 – Oct 2026

Distributed ETL Platform

- Developed a distributed RabbitMQ/MySQL-based ETL platform with a custom orchestration layer across **300+ machines** for large-scale modem log processing, enabling modem teams to support next-gen 6G development.
- Designed an **Integer Linear Programming (ILP)**-based compute allocation engine using Google OR-Tools to optimize machine assignments across queues based on pending workloads, machine capabilities, and capacity constraints, with queue-depth autoscaling to absorb bursty load.
- Re-architected modem log discovery from a polling-based crawler to an event-driven pipeline using SMB Change Notify, scaling to **300K events per day** and cutting log registration latency from **3 days to 1 minute**.
- Engineered a centralized observability platform using OpenTelemetry, Prometheus, Grafana, and Loki to provide pipeline metrics and log visibility, reducing Mean Time to Recovery (MTTR) by **30%**.
- Implemented **exactly-once** task processing using Redis TTL-based leases that workers acquire and renew via heartbeats, eliminating duplicate output rows and wasted compute while enabling automatic job failover to healthy workers.
- **Key Results:** Increased data processing capacity **30x** (1 TB/day to 30 TB/day), achieved **3x** faster data aggregation, and contributed to **$4,000** in compute savings.

Platform Engineering

- Designed a self-managed, highly available Kubernetes platform using lightweight k3s and Rancher, with a **3-master** control plane and **40+ worker nodes**, allowing the distributed data-processing workloads to run efficiently on existing compute
- Built GitHub Actions CI/CD pipelines with Ansible-based SSH rollouts to build, version, and deploy services across the machine fleet, replacing manual per-machine deployments with automated, repeatable releases.
- Built a centralized MySQL connection pooling service using Ray and SQLAlchemy, exposing query and update operations over gRPC to **100+ workers** to prevent database connection exhaustion, reducing backend MySQL connections by **50%**.

Associate Software Engineer – Infra Performance, Qualcomm, Bangalore Jun 2023 - Dec 2024

Automated Performance Testing Platform

- Built an automated hardware performance testing platform using RabbitMQ and Celery, with a custom DAG (Directed Acyclic Graph) based orchestration engine for dependency resolution across data collection, processing, and visualization jobs, replacing Excel workflows and ad hoc scripts. Orchestrated **2,000+ tasks/month** across **13+ devices**, improving overall efficiency by **83%**.
- Designed a central device-sharing service letting users discover, reserve, and connect to lab devices across multiple host machines from their laptops, tunneling ADB over HTTPS with authenticated access, exclusive locking, and auto-reconnect; **doubled** device utilization and reduced physical devices needed.

INTERNSHIP EXPERIENCE

Software Engineering Intern – Infra Performance, Qualcomm, Bangalore May 2022 – May 2023

- Engineered a polyglot persistence architecture (MySQL, MongoDB) for storage, post-processing, and analysis of memory performance benchmark results.

Android Developer Intern – GitHub Externship, FamPay, Remote Jan 2022 – Apr 2022

- Proposed and prototyped server-side UI rendering for Android applications (proposal), enabling UI changes to be delivered without app releases.
- Developed ArchyCards, an interactive arc-based advertising interface not natively supported by Android, applying computational geometry and custom UI rendering to deliver dynamic layouts with smooth performance.

RESEARCH EXPERIENCE

Explainable Multi-Disease Diagnosis and Mortality Prediction Apr 2026 – Present

Prof. Manish Kumar, VIT Vellore *Vellore, India*

- Reviewed **50+ research papers** to identify gaps in multi-disease diagnosis, mortality prediction, and explainable AI, informing the design of a unified machine-learning pipeline.
- Developed a three-stage ML pipeline using **36K+ NHANES records** to predict four diseases and five-year mortality, achieving ROC-AUCs of **0.94 for diabetes** and **0.93 for mortality**; applied **SHAP and DiCE** to explain predictions and generate counterfactuals for approximately 70% of high-risk cases.

PROJECTS

Apptitude: Hackathon Management Platform | GitHub Nov 2021 – Dec 2021

- Led backend development of a registration, team-formation, and OAuth authentication platform for an ACM hackathon with 1000+ participants using FastAPI and Firebase Realtime Database, deployed on Heroku.

Gleak: IoT Gas Leak Detection System | GitHub Nov 2021 – Dec 2021

- Built an end-to-end IoT gas leak detector: an ESP8266 and MQ-2 sensor stream readings to Firebase Realtime Database, a FastAPI server listens for leaks and sends push notifications, and an iOS app pairs with devices via a serial code to alert users.

EmoChat: On-Device Emotion-Aware Messaging | GitHub Aug 2020 – Nov 2020

- Developed a privacy-preserving Android chat app that translates facial emotion into live emojis, enabling expressive communication without exposing users' identities, using OpenCV and a TensorFlow Lite CNN for on-device inference.

Wi-Fi Controlled Street-Cleaning Robot | GitHub Aug 2019 – Sep 2019

- Prototyped a street-cleaning robot on an Arduino and ESP8266 NodeMCU, with a custom Android app sending control commands over Wi-Fi via TCP/IP, integrating embedded firmware with a mobile client.

TECHNICAL SKILLS

Languages: Python, Java, Kotlin, C++, SQL
Distributed Systems & Backend: gRPC, FastAPI, RabbitMQ, Celery, Kafka, Redis, Ray, Apache Spark
Infrastructure & Cloud: Kubernetes (k3s, Rancher, Helm, KEDA), Docker, Linux, AWS (EKS, EC2, S3, Lambda, Glue, Athena, SageMaker), GitHub Actions, Ansible, Jenkins
Databases: MySQL, PostgreSQL, MongoDB, DynamoDB, LanceDB, Apache Hudi, Apache Iceberg
Observability: OpenTelemetry, Prometheus, Grafana, Loki, Datadog, Splunk
AI/ML: LangChain, Agentic Systems (ReAct, Reflexion), Graph RAG, XGBoost, SHAP, TensorFlow Lite

LEADERSHIP

Tech Lead, Association for Computing Machinery (ACM), VIT Vellore chapter

- Mentored a team of 5 in App Development. Spearheaded the platform development for ACM Hackathon with 1000+ participants, enabling user registration and team formation

ACHIEVEMENTS

GitHub Externship, Winter Cohort

- Proposal for server-side UI rendering in Android apps selected from 8000+ submissions; implemented it during an industry externship at FamPay

Qualcomm All India Technical Conference (QBuzz)

- Presented the end-to-end automated performance testing platform built for the Infra Performance team; selected as one of 30 finalists from 500+ project submissions

Competitive Programming

- Top 10% worldwide in LeetCode contests, with 600+ problems solved across data structures and algorithms