

ANIRUDH RAGHAVENDRA MAKULURI

602-880-5334 | anirudh.makuluri@gmail.com | www.anirudh-makuluri.xyz | linkedin.com/in/anirudh-makuluri/ | Tempe, AZ, USA

SUMMARY

Full-stack engineer and AI/DevOps builder with 1.5+ years of industry experience building real-time, cloud-hosted systems for 12K+ users using React/React Native, Node.js, and AWS/GCP. I specialize in turning manual workflows into reliable, automated platforms, including LLM-powered tools and RAG systems that improve performance and developer velocity. Eligible for full-time roles in the US under OPT/STEM OPT starting May 2026.

EDUCATION

M.S Computer Science

August 2024 – Expected May 2026

Arizona State University, Tempe, AZ

GPA: 3.96/4.00

Relevant Coursework: Artificial Intelligence, Data Processing at Scale, Data Visualization, Mobile Computing, Data Mining

B.E. Electronics & Instrumentation

August 2018 – May 2022

Birla Institute of Technology and Science Pilani, Hyderabad, India

TECHNICAL SKILLS & CERTIFICATIONS

Languages: JavaScript/TypeScript, Python, Java, C#/C++, Kotlin, SQL

Frontend: React/React Native, Next.js, Vite, Zustand, TanStack Query, Tailwind CSS, Redux

Backend/APIs: Node.js/Express, FastAPI, REST/GraphQL, Socket.IO/WebRTC, Microservices, OAuth/JWT

Databases: PostgreSQL/Supabase, Neo4j, MongoDB, Redis, Firestore

Cloud/DevOps: AWS SDK (EC2/RDS), GCP, Docker Compose, GitHub Actions/CI-CD, SSM/Security Groups

AI/ML: TensorFlow, RAG/sentence-transformers, Scikit-Learn, NLP/Knowledge Graphs, LangChain, LangGraph

Certs: Oracle GenAI Professional (Aug 2025), DeepLearning.AI ML Specialization (May 2024)

PROFESSIONAL EXPERIENCE

Shardings — Software Engineer

June 2022 – August 2023

- Early-stage engineer owning architecture for a full-stack video conferencing platform (React, React Native, Node.js, Firebase) serving 12K+ users across web and mobile, driving feature parity so new capabilities shipped to both platforms in the same release cycle.
- Improved perceived frontend performance by integrating Google Cloud CDN and tuning image loading strategies, reducing median image load latency from 3 seconds to 800 ms and lowering bounce on media-heavy views.
- Implemented CI/CD pipelines using GitHub Actions, Fastlane, and Xcode Cloud, enabling zero-downtime mobile/web deployments, reducing deployment time from 45+ minutes to under 10 minutes, and achieving 99%+ success rates for 30+ prod releases.
- Built blue-green deployments for mediasoup SFU servers (separate per version): used PipeTransports to bridge old/new instances, enabling seamless cross-version media relay for users with zero downtime.
- Integrated Google Cloud Speech-to-Text and ChatGPT API to automatically transcribe and summarize live standups, eliminating 5 hours/week of manual note-taking per team and making recaps searchable for later audits.
- Mentored 5+ junior engineers and interns, conducted 20+ technical interviews, and used code/design reviews and pair programming to raise code quality and establish infrastructure best practices.

Shardings — Software Engineer Intern

July 2021 – December 2021

- Implemented full payment gateway with Razorpay and Stripe integrations supporting 3 pricing tiers (Basic, Standard, Pro); designed secure webhook handling, idempotency, and error recovery for reliable transactions.
- Developed a Google Workspace (Calendar) Add-on with JavaScript and Google Apps Script, integrating external calendar services and building a real-time room/resource scheduling interface with seamless data synchronization.

PROJECTS

SmartDeploy — GitHub-to-Cloud Deployment Platform

July 2025 – Present

- Built GitHub to AWS/GCP deployer handling 10+ services; Next.js dashboard reduced load times from 5s to under 1s using TanStack Query caching and Zustand state management.
- Built responsive UI components (RepoPageClient, DeploySheet, ScanProgress, Services dashboard) and kept deploy logs available across refreshes via server-side persistence.
- Enhanced DeploySheet with Vercel-backed custom URL checks and smart suggestions, turning invalid or taken domains into one-click alternatives so users never hit a dead end before deploying.

SD-Artifacts – LLM Repo Analyzer

Feb 2026 – Present

- Built 6-node LangGraph workflow on Amazon Bedrock (Claude Haiku), analyzing GitHub repos to generate production Dockerfiles, docker-compose, and nginx configs; 85% confidence across Node/Python stacks powering SmartDeploy's infra generation.
- Developed FastAPI service with /analyze/stream SSE endpoint streaming progress across scanner, planner, generator, and verifier nodes for realtime dashboard integration in SmartDeploy.
- Scanner parsed GitHub API file tree (package.json, requirements.txt) to detect stack, services, ports; returned structured JSON with risks, hadolint results, and confidence scores.

Accio — PDF RAG Knowledge System

Sept 2025 – Dec 2025

- Built a PDF RAG system (FastAPI, Streamlit) that converted 2.8K pages of federal documents into a Neo4j knowledge graph with 3K+ nodes and 4.9K relationships, achieving 77.5% F1 on entity extraction.
- Designed a layout-aware extraction pipeline using Docling, PyMuPDF, and Gemini to feed local MPNet embeddings, improving retrieval quality by 25% over baselines (NDCG@5).
- Optimized end-to-end processing to 1.15 minutes per document and Dockerized Neo4j to handle 1K+ text chunks while preserving context.