

ANIRUDH RAGHAVENDRA MAKULURI

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

SUMMARY

Adept at converting manual workflows into reliable systems, recently with agentic deployment tools, RAG-based chat, and real-time platforms on AWS/GCP. Currently exploring how LLMs, observability, and automation can make full-stack teams ship faster with less operational overhead. Eligible for full-time employment in the US for up to 36 months under OPT/STEM OPT beginning May 2026.

EDUCATION

M.S Computer Science <i>Arizona State University, Tempe, AZ</i> Relevant Coursework: Artificial Intelligence, Data Processing at Scale, Data Visualization, Mobile Computing, Data Mining	August 2024 – Expected May 2026 3.96 GPA
B.E. Electronics & Instrumentation <i>Birla Institute of Technology and Science Pilani, Hyderabad, India</i>	August 2018 – May 2022 3.2 GPA

TECHNICAL SKILLS & CERTIFICATIONS

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

Frontend: React, React Native, Next.js, Redux, HTML, CSS, Tailwind

Backend & Databases: Node.js, Express.js, REST APIs, PostgreSQL, NoSQL (Firebase Firestore, MongoDB), MSSQL

Cloud & DevOps: GCP (Cloud Run, CDN, Firestore), AWS (EC2, RDS, ECS Fargate, Elastic Beanstalk, Amplify), Docker, Kubernetes

Machine Learning & AI: TensorFlow, PyTorch, Scikit-Learn, LangChain, Prompt Engineering, RAG, LLMs (Gemini, Bedrock, Mistral), NLP

Tools & Others: Git, GitHub, Figma, Agile/Scrum, OAuth, Linux/Unix

Certifications: Oracle Certified Generative AI Professional (August 2025), Machine Learning Specialization (May 2024)

PROFESSIONAL EXPERIENCE

Shardings: Software Engineer

June 2022 — August 2023

- Architected full-stack video conferencing platform (React, React Native, Node.js, Firebase) serving 12K+ users across web and mobile, and drove feature parity so new capabilities shipped to web and mobile in the same release cycle.
- Designed and operated on-call workflows with blue-green deployments, real-time dashboards, and admin/crash-log tooling, cutting incident investigation and recovery time from 45+ minutes to under 15 minutes (67%).
- Improved perceived performance for 12K+ users by integrating Google Cloud CDN with cache invalidation, reducing median image load latency from 3 seconds to 800 ms and lowering bounce on media-heavy views.
- Integrated ChatGPT with Google Cloud Speech-to-Text to auto-transcribe and summarize live standups, eliminating 5 hours/week of manual note-taking per team and making recaps searchable for later audits.
- Established CI/CD pipelines with GitHub Actions (build, deploy, and automated rollback on error thresholds), achieving a 99%+ successful deployment rate and enabling multiple safe releases per week.
- Mentored 5+ junior engineers and interviewed internship candidates, using pair programming, code and design reviews to raise code quality.
- Led the mobile app from prototype to production-ready, owning navigation, home/meetings/recaps flows, and release pipelines so mobile usage could scale alongside the web client without a separate platform team.

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 Google Workspace Add-on using JavaScript and Google Apps Script APIs to integrate with external calendar services; built real-time room/resource scheduling interface with seamless data synchronization.

PROJECTS

SmartDeploy — Agentic DevOps Platform

July 2025 – Present

- Built agentic deployment platform (Next.js, Node.js) with AI-driven orchestration across AWS/GCP; enabled one-click deploys for 10+ repos without manual infra.
- Built a Gemini-powered code-analysis pipeline that infers build and deployment configurations for dozens of heterogeneous GitHub repositories, enabling new services to be onboarded without writing configuration from scratch.
- Added dynamic target selection from code analysis (deps, concurrency, DB needs); auto-provisions RDS for stateful apps.
- Built deployment telemetry layer (structured logs, deploy events, health checks) feeding a single dashboard, so engineers can see rollout status and failures across AWS and GCP targets in real time and debug issues faster.

Chatify — End-to-End Encrypted AI Chat Platform

December 2024 – Present

- Built real-time chat app (Socket.IO, React/React Native) with 1:1/group messaging, media sharing, timezone-aware scheduling, and infinite scroll; tuned it for sub-second delivery under typical load.
- Integrated Gemini 2.5 Flash API for a personalized AI assistant and used Zep to persist long-term conversation state so responses stay consistent across sessions.
- Implemented semantic search over chat history using Hugging Face sentence embeddings (all-MiniLM-L6-v2), allowing users to instantly retrieve relevant past messages.
- Implemented message-level end-to-end encryption using libsodium on top of TLS-secured WebSocket (wss) connections so only intended participants can decrypt chat content