Abhinav Kumar Mallick

Senior Software Engineer — Distributed Systems — AI Systems

+91-7004297271 | abmallick101@gmail.com | linkedin.com/in/abml | github.com/abmallick

SUMMARY

Senior Software Engineer with 5+ years of experience building and scaling real-time data-heavy systems. Experience with AI agents and voice AI systems, alongside deep expertise in AWS and zero-downtime infrastructure.

EXPERIENCE

Dreamsleep Healthcare Pvt Ltd   Aug 2024 – Sep 2025
Founding Software Engineer Lead (SDE III)   *Bengaluru, India*

- Led architecture and implementation of a distributed backend platform supporting **10k+ IoT devices**, designing APIs, data models, and workflows for real-time telemetry ingestion and analytics.
- Designed real-time ingestion pipelines processing high-frequency sensor data, enabling near-real-time ML inference while maintaining **p99 latency under 200ms under peak load.**
- Scaled backend systems to support **10,000+ concurrent devices and 10k RPS** user traffic, introducing load testing, backpressure controls, and performance benchmarking to validate system limits.
- Owned production reliability through **zero-downtime deployments**, CI/CD automation, and observability pipelines (metrics, tracing, alerts), reducing operational risk during rapid iteration.
- Mentored engineers through **regular 1:1s**, defined performance metrics, and guided technical growth.

Groky   Feb 2024 – Jun 2024
Founder   *Bengaluru, India*

- Built an AI-assisted developer platform for large codebases, focusing on program analysis, context retrieval, and latency-sensitive inference.
- Conducted structured evaluations with **30+ engineers** to iterate on retrieval quality and UX, and fine-tuned LLMs for code understanding workflows.

Nuclei   Jul 2020 – Feb 2024
Software Development Engineer II   *Bengaluru, India*

- Led the central engineering team driving performance optimization, architectural reviews, and reliability patterns across **multiple product lines.**
- Built an internal **RAG-based knowledge chatbot** for marketing and sales teams, enabling instant technical Q&A over large internal documentation and product data using embeddings and retrieval pipelines.
- Designed and scaled a high-throughput analytics platform using ClickHouse, reducing query latency by **100x** and infrastructure cost by **90%**, enabling near-real-time analytics for multiple product teams.

ACHIEVEMENTS

Open Source: Published datasets for network traffic anomaly and sleep-stage biometrics. Hugging Face
Research: Published papers on AI for security and IoT data protection. Google Scholar
Awards: KPIT Innovation Award (Smart India Hackathon 2018) and INR 2.75L grant from NAIN, Govt. of Karnataka.
Certifications: Completed DeepLearning.AI Deep Learning and TensorFlow Developer specialisations.

TECHNICAL SKILLS

Languages: Go, Python, SQL
AI Systems: Voice AI Agent, RAG, Multi agent system and workflows
Distributed Systems: gRPC, REST, Kafka, ClickHouse, Redis, MQTT; event-driven architectures, caching, consistency models, circuit breakers, fault tolerance, load balancing
Cloud & Infra: AWS (EC2, EKS, Lambda, SQS, EventBridge, SageMaker), Kubernetes, Docker, Terraform, CI/CD
Observability: Prometheus, Grafana, metrics, tracing, alerting, performance benchmarking
Core Concepts: Data structures and algorithms, concurrency, system design, API design, scalability, reliability

EDUCATION

Siddaganga Institute of Technology   Tumkur, Karnataka, India
Bachelor of Engineering - 8.7 GPA   *Aug 2016 – May 2020*