

ABHIJITH GANESH

Chennai, India

✉ work@abhijithganesh.com

🌐 [linkedin.com/in/AbhijithGanesh14](https://www.linkedin.com/in/AbhijithGanesh14)

🐙 github.com/AbhijithGanesh

🌐 AbhijithGanesh.com

Founding AI engineer shipping production LLM systems end-to-end – retrieval pipelines, document forensics, and agentic automation – on cloud-native infrastructure. CKA-certified, with measurable wins in latency, build speed, and reliability.

Experience

Exaqube

Apr 2025 – Present

Founding AI Engineer

Chennai, India

- Built TamperCheck, a document forensics service detecting tampering in user-uploaded PDFs and images, integrated into the core ingestion pipeline.
- Delivered the CharterQube intelligence platform end-to-end – LLM-backed retrieval over private corpora with grounded citations.
- Led end-to-end customer-facing product development from requirements to production across multiple AI-driven workflows.
- Standardized CI/CD across **90+ repositories** (GitHub Actions, Docker, GHCR) – cut CI failures by **60–80%** and builds **50–70%** faster.
- Reduced Azure spend by consolidating workloads onto fewer VMs and leasing out idle compute capacity, materially lowering monthly cloud cost.
- Architected multi-cloud failover and led the weekend incident response that restored **19 Azure VMs** with zero data loss.
- Migrated **35+ production applications** to a React Router v7 + Cloudflare edge standard with zero downtime across the rollout.

Peritys

Jan 2025 – Apr 2025

Backend Developer

Chennai, India

- Implemented Redis caching, cutting API response times from **1500ms** to **200ms**.
- Reduced Docker build times from **20 min** to under **5 min** via layer ordering and dependency caching.
- Replaced a dedicated upload microservice with presigned Bunny CDN uploads, eliminating a service from the topology.
- Shipped real-time presence features (“Who Just Joined”, “Who’s Online”) with WebSocket-backed broadcast.

Projects

Project Agentic | Gemini, FastAPI, Office API

- LLM-driven agent automating machine-level tasks (mailing, productivity workflows) with remote multi-language code execution (JS/TS, Python, Java, C++).

Rekognize | PyTorch, FaceNet, MTCNN, FastAPI



- Attendance system using MTCNN face detection + FaceNet embeddings for individual identification.

Certifications

Microsoft Certified: Azure AI Fundamentals (AI-900), Azure Fundamentals (AZ-900)

December 2023

Cloud Native Computing Foundation: Certified Kubernetes Administrator (CKA)

August 2023

Amazon Web Services: AWS Certified Practitioner

December 2023

Google Cloud Platform: Cloud Digital Leader

January 2024

Technical Skills

AI/ML & Deep Learning: PyTorch, TensorFlow, HuggingFace, OpenCV, FaceNet, MTCNN, CometML, Ollama

LLMs & Retrieval: OpenAI APIs, Google Gemini, Qdrant, RAG pipelines, prompt evaluation

Languages: TypeScript, Go, Python, SQL, C/C++, Java

Frameworks: FastAPI, Django, Next.js, React 19, Gin, Fiber, Astro

Databases: PostgreSQL, Redis, MongoDB, Cassandra, Cockroach, MySQL

Cloud & Infra: Azure (AKS, AzureML, DevOps), AWS (EC2, S3, SageMaker), Kubernetes (CKA), Docker, Cloudflare, Prometheus, Grafana, Kafka

Education

Vellore Institute of Technology, Chennai

Sep 2021 – Jun 2025

B.Tech, Computer Science (AI & Robotics) – CGPA 8.95

Chennai, India

Indian Institute of Technology, Madras

Aug 2021 – May 2025

BS, Data Science & Programming (Online Degree)

Chennai, India