Sangeethkumar

AI Engineering Leader · Engineering Manager, AI Transformation & Platform Engineering

Chennai, India · open to remote
sangeethcloud@gmail.com +91 8220204407 linkedin.com/in/sangeethk22 github.com/sangeethkumar22 me.shipscale.org

Profile

AI engineering leader and engineering manager who builds production systems **and validates the measurements behind them**. Eleven years in enterprise SaaS, the last two and a half managing engineering teams – currently an **engineering function of 30+ engineers** – and hands-on building production AI end to end: retrieval, evaluation, and the governance layer deciding what reaches users. I'm leading the function's move from traditional delivery toward AI-native execution, and measuring honestly how far it has got, while owning reliability, capacity and delivery outcomes. I pair system design with the discipline to scale teams, grow the people in them, and turn engineering reality into something leadership can decide on. My distinguishing habit is the unglamorous one: **checking the ruler before trusting the reading**.

Focus areas

Context engineering · LLM evaluation · Production RAG · Agentic systems · Multi-agent orchestration · AI governance · AI observability · Model routing · LLMOps · Contract-driven development · AI-native SDLC · Platform engineering · Platform-as-product · Distributed systems · Reliability & incident RCA · Enterprise readiness · Engineering management · Capacity & headcount modelling · Talent retention · Developer productivity · Executive communication

Experience

Kissflow · Orangescape Technologies

Engineering Manager – AI Transformation & Platform Engineering

Apr 2024 — Present

- Lead an engineering function of **30+ engineers**, owning delivery, reliability and the growth of the people in it. Led the function through a period of turnover.
- Leading the function's move to AI-native execution across architecture, development, QA, debugging and operations without mandates, and measuring where individual gains stop short of organisational ones.
- Personally built and shipped **two production AI systems** – the knowledge assistant and the work-progress platform – while running the function, and led a configuration-analysis service into a **live customer approval path**, called when a request is submitted and before it is approved.
- Designed a governance model for the context feeding an LLM: automated jobs build candidates only and never write to live, a contract gate blocks unsafe changes, humans promote, pinned versions roll back in one click. Made candidate-only a hard rule after an ungoverned nightly sync cut a production index by **more than three-quarters**.
- Established evaluation-first development – then found the **evaluation itself was unreliable**, with enough run-to-run variance that every apparent gain in an optimisation sweep sat inside the noise band. Moved decisions to offline replay over already-judged answers, confirmed with multi-run significance testing.
- Built a **capacity and demand model** that turned "the team feels underwater" into a month-by-month backlog projection for a leadership briefing – planned against effective capacity rather than headcount, because those are different numbers. **Then scored the forecast against actuals**, found it over-pessimistic, and traced why.
- Led a platform-wide enterprise-readiness assessment across dozens of services, producing an **executive-ready reliability roadmap**, and root-caused a severe production CPU-spike incident.
- Initiated and led a **company-wide AI hackathon** – ten days, real product and engineering problems, teams asked to deliver working solutions – deliberately aimed past code generation at architecture thinking, debugging and problem-solving. Set review expectations for AI-generated code: mandatory self-review, TDD-first, design review before implementation.
- Orchestrated a cloud-storage cost optimisation, delivered through the team, worth a **five-figure annual** saving.

Lead Software Engineer – Workflow Management

Dec 2021 — Apr 2024

- Owned the Workflow Management product end to end – **data modelling**, workflow automation, technical direction, and analytics delivered directly on the application platform.
- Worked on complex MongoDB queries behind the product's analytics.
- Led the team's technical decisions.

Lead Software Engineer – Case Management

Feb 2020 — Dec 2021

- Built the Case Management product **0→1** – product analysis, data model, system design, **real-time sync**, and the state machine that moves every case through its lifecycle – then led the team's technical decisions and design reviews.
- Used Google Firebase for real-time event sync in the product.
- Led the team and worked on its technically complex features.

Senior Software Engineer – Search, Notifications & Audit

May 2018 — Feb 2020

- Built the platform's complete audit system and per-user activity feed; owned search and notification delivery across the product.

- Worked on search with Elasticsearch and notification delivery with AWS SES.

Corestack

Software Engineer Apr 2016 — May 2018

- Cost and compliance optimization and recommendation for a multi-cloud governance platform.

  Python · MongoDB · AWS

CD Cloudenablers

Junior Software Associate – Infrastructure Fulfillment Oct 2014 — Apr 2016

- Built an automated infrastructure-approval portal for multi-cloud provisioning from scratch.
- Brought multiple cloud infrastructures into one portal, with an automated approval process for infrastructure requests from IT teams.
- Worked on Node.js as a backend developer.

  Node.js · MongoDB

Selected systems

AI Knowledge Assistant

Designed and built end to end – an internal assistant giving engineers source-grounded answers about a large multi-service platform. Agentic retrieval with tool use, a knowledge pipeline that keeps its knowledge current, and a category-balanced evaluation harness scoring correctness, completeness and faithfulness separately. **Made answer quality a measured number rather than a feeling** – and the honest number held back a wider rollout.

Python · FastAPI · Next.js · PostgreSQL · pgvector · Cloud Run · Claude / GPT / Gemini

Context Governance Framework

A lifecycle for the knowledge feeding an LLM: candidate-only builds meant to keep automated jobs off live, a contract gate that blocks silent shrinkage and lost facts (answer scoring advisory), manual reversible promotion, pinned versions, one-click rollback. Built after unattended automation shrank a production index overnight, and tightened after a later run gutted it. Includes the uncomfortable finding that the governance layer was **built but inert** – every field present, the authority ranking a no-op – because provenance was never populated.

Python · PostgreSQL · Cloud Run · scheduled jobs

AI-Native SDLC – Multi-Agent Engineering Workflow

Designed a software-delivery workflow built from specialised agents – product, architecture, development, QA, security, DevOps and an orchestrator – coordinated by **contract-driven development**: architecture contracts are frozen before implementation starts, changes pass through structured governance, and a human approves contracts, foundational changes and final closure. Paired with review expectations for AI-generated code: mandatory self-review, TDD-first, and design review before implementation.

Claude · OpenAI · Gemini · agent orchestration · persistent session state

Configuration Analysis Service

Led development and rollout of a configuration-analysis service – **health scoring, dependency-impact analysis** and optimization recommendations, with stakeholder reporting – **invoked inside a live customer approval flow**.

React · Vite · FastAPI · MongoDB · Cloud Run

Work Progress Intelligence Platform

Sole architect and builder of an internal platform: conversational AI standups that extract structured updates, two-way sync with the work-tracking system, and dashboards shaped to each role. Raising its chat assistant's high-confidence answers from 81 to more than 90 out of 100 (the assistant's own confidence label) was not a data problem: the data was always there, and the failures were routing gaps and tool bugs.

Next.js · TypeScript · FastAPI · PostgreSQL · Redis · Celery

Engineering Capacity & Demand System

Models demand against **effective capacity** rather than headcount and projects the backlog breakpoint as staffing changes, feeding a leadership briefing. **Then I scored my own forecast against actuals and it was wrong** – at two months, the bug projection by more than twice over, the broader forecast by about 30%. The bug backlog had improved because fewer bugs came in, not because throughput rose.

Python · REST integrations · scheduled refresh · executive reporting

Enterprise Readiness Assessment

Led a platform-wide assessment across dozens of services – examining service behaviour, production telemetry and incident history through a structured multi-perspective review. Produced an **executive-ready roadmap** for performance, reliability and cost at scale.

MongoDB · GCP · Kubernetes · Snowflake · BigQuery

Analytics Pipeline Reliability Redesign

Led the reliability audit of an analytics data pipeline and authored a resilient redesign — self-correcting replication and automatic reconciliation — with a migration and operations plan.

MongoDB · Pub/Sub · Snowflake · GCP · change data capture

Customer Stability Dashboard & Incident RCA

Designed and built a per-customer health view joining production telemetry with support load — closing the gap where engineering and support each held half the picture. Separately root-caused a **severe production CPU-spike incident** following an infrastructure upgrade and wrote up the fix.

FastAPI · BigQuery · MongoDB · Kubernetes

Skills

AI & LLM engineering	Evaluation-first development · eval harness design · LLM-as-judge · production retrieval systems · hybrid search and reranking · multi-agent orchestration · tool-calling and routing · prompt versioning · Claude, GPT, Gemini
AI governance & reliability	Runtime governance · contract gates · human-in-the-loop promotion · audited rollback · provenance and lineage · least-privilege agent tooling · faithfulness and abstention policy · token-consumption monitoring · AI observability · model routing
Leadership & architecture	Engineering management · team scaling · AI-native SDLC · contract-driven development · multi-agent engineering workflows · platform-as-product · system design · data modelling · architecture impact analysis · distributed systems · multi-tenant SaaS · tenant isolation · capacity modelling · talent development and retention · enterprise readiness · developer productivity · peer code review
Backend & languages	Python · FastAPI · Node.js · TypeScript · SQL · asynchronous services · API design · Celery
Data & storage	PostgreSQL · MongoDB · Snowflake · BigQuery · Elasticsearch · Redis · Firebase (real-time sync)
Cloud & infrastructure	Google Cloud (Cloud Run, Pub/Sub, GCS) · AWS (API Gateway, Lambda, SES, SNS, SQS, WorkMail) · Kubernetes · Docker · CI/CD · Jenkins (Groovy scripts for continuous deployment) · security testing with OWASP ZAP · observability · incident RCA
Frontend	React · Next.js · TypeScript · Vite · Tailwind

Education

B.E. Electronics & Communication Engineering — 2011 — 2014

Mepco Schlenk Engineering College, Sivakasi

Diploma, Electronics & Communication — 2008 — 2011

Alagappa Polytechnic College, Karaikudi