

ALI RATHORE

AGENTIC SYSTEMS · AI
PLATFORMS · DATA
INFRASTRUCTURE

Founding engineer and engineering leader at the convergence of data infrastructure and AI. I lead architecture and engineering for an agentic data fabric platform: ingestion, storage, transformation, semantics, and serving on the data side; retrieval, agentic systems, evaluation, and ML inference on the AI side. Over fifteen years I have built platforms and the teams behind them: database engines at SAP, a data platform Oracle acquired, engineering organizations across healthcare, and distributed senior teams I founded and led across three continents.

CONTACT

ali1rathore@gmail.com
San Francisco Bay Area, CA
linkedin.com/in/a8e

SKILLS

Agentic Systems, RAG, LLM
Evaluation, ML Serving
Data Infrastructure, Distributed
Systems, Kubernetes
Engineering Leadership,
Technical Hiring, Build-vs-Buy
Founder, Mentor, Teacher

EDUCATION

B.S. ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

UC Berkeley | 2010
Summa Cum Laude · Haas
Scholar

WORK EXPERIENCE

FOUNDING ENGINEER

2025 – Present

Fabrion · Agentic data fabric platform · San Francisco Bay Area

- Own the platform's architecture end to end: API contracts, identity model, agent execution, data architecture, and Kubernetes delivery for a platform that lets enterprises run AI agents safely against their operational data.
- Defined the contract-first architecture that turns the platform's open-source data and AI subsystems into one governed product: a single identity flowing through every layer, so subsystems stay swappable and every agent action is attributable to a user.
- Led agent execution from design through production: workspace isolation, session recoverability, and policy-governed tool access (MCP) that holds up to enterprise security review.
- Set technical direction for document intelligence: an ingestion and retrieval architecture (multi-modal RAG) that turns heterogeneous document corpora into queryable knowledge, with every answer traceable to its source page, which customers required before trusting AI-generated answers.
- Made the platform installable and operable as a product. The full 15-service stack deploys into a customer's Kubernetes environment in one step, with tenant isolation, single sign-on across every component, and a release-gating integration suite.
- Hire and ramp the platform's engineers, keep every architecture decision written down with the options rejected and why, and run the build-vs-buy evaluation behind each major subsystem.

FOUNDER, CTO

2023 – 2025

Lakeshore Labs · AI agents and automation consultancy (SEA and GCC) · Remote

- Designed and shipped production agent systems for enterprises in F&B, insurance, distribution, and industrial IoT. The recurring work: grounding answers in verifiable sources, evaluation that catches regressions before customers do, and integration with the systems of record enterprises run on.
- Built a research agent for a major F&B group that synthesizes trend and product research from 100+ sources into briefs its innovation team acts on directly, with source weighting tuned by the team's own feedback.
- Built a benefits-and-claims assistant for a health-insurance provider: identity-scoped retrieval over plan documents and claims systems, so members resolve questions conversationally instead of through the call center.
- Built a natural-language interface to distributed IoT fleets that lets non-technical operators query and control devices safely: permissioned command paths, blast-radius limits, and confirmation on destructive operations.

SENIOR ENGINEERING MANAGER

2022 – 2024

Soda Health · Health-equity fintech · Chicago, IL (Hybrid)

- Owned member-experience engineering: the AndMore mobile app, web portal, engagement, customer segmentation, recommendations, and LLM integration.
- Took the company mobile, launching iOS and Android apps in 3 months with a team hired and ramped in parallel, opening a new revenue channel.
- Shipped an in-app AI assistant answering members' benefits questions in real time, barcode scanning built on a custom vision model, and a geolocation recommendation engine across hundreds of partner stores.

ALI RATHORE

AGENTIC SYSTEMS · AI
PLATFORMS · DATA
INFRASTRUCTURE

TECHNICAL SKILLS

Agent harnesses, MCP, tool use,
sandboxing, sub-agents

RAG and hybrid retrieval, LLM
evaluation, fine-tuning

Claude, OpenAI, Gemini APIs,
LangChain, LangSmith

Ray, vLLM, KubeRay inference
serving

Go, Python, TypeScript, C++,
SQL, Java

Postgres/pgvector, ClickHouse,
Milvus, Neo4j

Spark, Kafka, BigQuery, Dagster,
Airbyte, Cube.js, OpenMetadata

Kubernetes, Helm, K8s
operators, Docker

AWS, GCP, Azure, OCI,
Keycloak/OIDC

GitHub Actions, GitLab CI,
OpenTelemetry

VOLUNTEERING

MAGNITUDEMINDS, VIETNAM

Chief Technical Advisor | 2025 –
Present

CODERSCHOOL, VIETNAM

Teacher, Blockchain | 2019 –
2021

UC BERKELEY, CA

Researcher, Nanowires | 2009 –
2010

ENGINEERS WITHOUT BORDERS

Project Manager | 2008 – 2009

WORK EXPERIENCE CONTINUED

ENGINEERING MANAGER

2021 – 2022

eVisit · Enterprise telehealth serving 1M+ patient interactions · Remote

- Led the 8-person product team shipping the enterprise suite (SSO, self-service onboarding, billing, permissions) that unblocked large health-system deals.
- Directed, in parallel, the 4-person infrastructure team's Azure-to-AWS Fargate migration, cutting infrastructure costs by \$100K+ annually.

DIRECTOR OF ENGINEERING

2019 – 2021

Metaprise Systems · AI, cloud, SaaS, and IoT applications studio · USA, Netherlands, Pakistan, Vietnam

- Led delivery of AI/ML client solutions across four countries while scaling a distributed team of 20+ engineers. The work spanned computer-vision dental shade matching (productized as Matisse.ai), IoT fleet management for a Philippine delivery network built on in-truck telemetry, and real-time transaction anomaly detection for a US convenience-store chain.

CONSULTING MEMBER OF TECHNICAL STAFF

2018 – 2020

Oracle · Redwood City, CA

- Guided 6 teams (33 engineers) to launch Spark-as-a-Service on Oracle Kubernetes Engine in 12 months; ran the Fortune 500 proofs of concept that shaped the platform for real-time analytics and ML workloads.

PRINCIPAL SOFTWARE ENGINEER

2016 – 2018

SparklineData · Acquired by Oracle · San Jose, CA

- Architected one of the first data lake platforms built on compiled query execution: interactive OLAP directly over open data formats on Apache Spark and Kubernetes, sustaining 10,000+ queries per second.

DIRECTOR OF SOFTWARE

2014 – 2016

Momentum Machines · Autonomous food robotics, now Creator · San Francisco, CA

- Led the 15-engineer team building the real-time control system, firmware, and computer vision for robotic burger assembly, with 20+ units in the field under centralized fleet oversight.

STAFF SOFTWARE ENGINEER

2011 – 2014

SAP · HANA in-memory database core · USA, Germany, Korea, France

- Engineered distributed query processing and in-memory vectorized execution in C++ inside the HANA core, including a lockless persistent hash table that improved parallel query throughput 200%, with hot paths tuned using SIMD and RDMA.
- Led HANA's Hadoop integration across engine teams in four countries, bringing external data into the distributed query planner.

SENIOR MANUFACTURING ENGINEER

2010

Solyndra · Thin-film solar manufacturing · Fremont, CA

- Built real-time process-monitoring software for thin-film solar cell production lines, instrumenting manufacturing steps and surfacing live yield and cycle-time metrics that drove the fab's process improvement initiatives.