

Ratnadeep Bhattacharyya

Senior Backend & AI Platform Engineer – Distributed Systems, LLMs & Cloud

✉ratnadeep.bhattacharyya@gmail.com ☎+91 8310823305 🌐https://theskylab.in/

Summary

Backend and AI Platform Engineer with 8 years of experience designing and operating distributed systems, high-scale APIs, event-driven architectures, and production LLM platforms. Hands-on experience across Node.js/TypeScript, Python, Go, Kubernetes, AWS/GCP, RAG, agentic workflows, and asynchronous processing. Built backend systems handling 1M+ daily API requests and AI workflows serving thousands of users, with end-to-end ownership spanning architecture, implementation, production reliability, observability, evaluation, and stakeholder collaboration.

Experience

Springboard (SlideRule Inc.) India (Remote)
Software Engineer II (Fullstack) Jan 2025 - Present

- Built and productionized a **multi-step agentic LLM grading system** combining orchestration, contextual retrieval, tool execution, stateful evaluation workflows, and asynchronous processing, reducing grading turnaround from 6 days to 2 hours with 93% accuracy across 5,000 students.
- Built and owned full-stack services and workflows for Springboard’s mentorship platform using **Python, Go, React, and MySQL**, automating student–mentor scheduling based on availability and reducing manual coordination.
- Built **AI call-summary workflows** that convert mentor-session transcripts into structured summaries, action items, next steps, and support-risk indicators, reducing manual follow-up for student support teams.
- Designed **distributed workflow orchestration** with retries, idempotency controls, fault tolerance, and operational traceability across asynchronous grading, mentorship, and support workflows.
- Owned architecture and technical decisions** across agent orchestration, backend services, asynchronous execution, evaluation, observability, and production reliability for AI-powered workflows.
- Built **golden evaluation datasets and regression checks** for agent outputs and tool usage, using production traces and failure cases to measure quality across model and workflow changes.
- Built **observability and LLM tracing** using LangSmith, Prometheus, Grafana, structured logging, distributed tracing, and New Relic to debug agent/tool execution and production failures.
- Deployed and operated workflow services on **Kubernetes (GKE)** with CI/CD pipelines, improving scalability, release reliability, and production availability.

TarkaLabs India (Remote)
Technical Consultant Jan 2024 - Nov 2024

- Worked directly with product stakeholders and end users to translate customer workflows and retrieval requirements into system architecture, technical decisions, and production capabilities for a **self-hosted agentic RAG platform**.
- Built and delivered a self-hosted **agentic RAG system** using Python, FastAPI, PostgreSQL, Qdrant, LangChain, and LlamaIndex for querying structured and unstructured enterprise knowledge.
- Designed retrieval and reasoning workflows using **vector search, hybrid search, reranking, prompt rewriting, contextual retrieval, NL-to-SQL, and NL-to-API** to improve relevance and automate customer-specific workflows.
- Implemented **RAGAS-based evaluation, PII filtering, prompt safety checks, output validation, and observability** to improve reliability and production quality across AI workflows.
- Built scalable **FastAPI backend services** for asynchronous ingestion, retrieval orchestration, external API integrations, and agentic workflow execution.
- Owned multimodal ingestion for **PDFs, Word documents, Excel files, images, tables, audio, and video transcripts**, enabling enterprise content to be processed for downstream RAG workflows.
- Developed custom chunking and metadata strategies to preserve **page references, table context, and source traceability**, improving citation quality and generated-answer trust.

Nagarro India (Remote)
Senior Software Engineer July 2023 - Dec 2023

- Improved reliability across gift and payment workflows handling **1M+ daily API requests** by redesigning service communication patterns, reducing cascading failures, and improving fault isolation.
- Built and optimized high-volume Node.js/TypeScript/NestJS** backend services integrated with payment and order systems using **MySQL, RabbitMQ, and Redis**.
- Applied **Saga-based distributed transaction patterns** with RabbitMQ-based asynchronous processing to improve consistency, recovery, and failure handling across multi-service workflows.
- Enhanced backend performance through **MySQL query optimization, Redis caching, and asynchronous processing** to support low-latency transactions under high traffic.
- Designed reliable purchase workflows coordinating gift issuance, payments, order state, and dependent microservices.
- Strengthened production debugging and operational visibility through structured logging and monitoring.

Dentsu World Services India (Remote)
Software Engineer July 2021 - June 2023

- Built **Python and Go microservices** for a multi-tenant analytics platform serving **6K+ daily users and 200K+ daily API requests** across telecom, manufacturing, and logistics customers.
- Built high-throughput **Go-based event ingestion systems** for distributed customer data streams used in analytics, reporting, and operational workflows.
- Owned design of **event-driven backend architectures and asynchronous processing pipelines**, balancing scalability, fault tolerance, service communication, and operational reliability.
- Developed AWS-based backend services using **ECS/EKS, Lambda, S3, and Airflow** for batch, near-real-time, and distributed processing workloads.
- Designed external-facing APIs and webhook workflows allowing third-party clients to submit analytics jobs, consume outputs, and receive asynchronous completion notifications.
- Designed ingestion and storage workflows using **MongoDB and Snowflake** for high-volume analytics processing.
- Improved throughput, cost efficiency, and reliability through **batching, caching, optimized data flows, retries, failure recovery, logging, and monitoring**.

Cura Technologies Inc Bangalore, IN
Software Engineer Dec 2019 - July 2021

- Built backend APIs and distributed services using **Python, Django, Flask, and PostgreSQL** for a healthtech platform serving **10+ hospitals and ~600 active users**.
- Led migration of backend services to **Kubernetes/EKS**, defining containerization, deployment, and scaling practices to improve reliability and operational consistency.
- Developed a **custom Kubernetes operator** to automate PostgreSQL provisioning, failover management, and lifecycle operations.
- Designed AWS infrastructure using **Terraform** and implemented rolling and blue-green deployments to improve scalability and release reliability.
- Built Flask-based report-generation microservices to automate clinical reporting and improve turnaround time for hospital operations.
- Built Flutter-based healthcare functionality, including a **DICOM interface**, and contributed to migration from Swift to Flutter across iOS, Android, and web.
- Worked directly with doctors, nurses, healthcare operations teams, and other stakeholders to translate clinical requirements into production improvements.

Tapchief (Relel) Bangalore, IN
Product Engineer Oct 2018 - Oct 2019

- Built and shipped end-to-end product features across **Python, Node.js, PostgreSQL, Angular, and React Native** for a platform serving **~30K daily active users and ~100K daily API requests**, owning workflows from backend APIs through frontend integration and production delivery.
- Developed **multi-step onboarding and real-time communication workflows** using Angular, WebRTC, and Twilio, supporting **~80–100 audio/video calls per day** across web and mobile products.
- Built internal operational tooling and improved **debugging, deployment stability, and production reliability** across backend and frontend systems.

Skills

Backend & Distributed Systems: Node.js, NestJS, Python, FastAPI, Django, Flask, Go, REST, gRPC, Microservices, Event-Driven Architecture, Distributed Systems, System Design, API Design

Languages: TypeScript, JavaScript, Python, Go

AI/LLM Systems: Multi-Agent Orchestration, Tool Calling, Agent Memory & State, Context Management, Agentic RAG, Hybrid Retrieval, Reranking, Structured Outputs, LLM Evaluation/RAGAS, Prompt Safety, PII Filtering

Data & Messaging: PostgreSQL, MySQL, MongoDB, Snowflake, BigQuery, Redis, Qdrant, RabbitMQ, Pub/Sub, Queues, Vector Search, Embeddings

Cloud & Infrastructure: AWS, EKS, Lambda, S3, GCP, GKE, Kubernetes, Docker, Terraform, CI/CD

Observability: LangSmith, Prometheus, Grafana, Distributed Tracing, Structured Logging, New Relic, Production Debugging

Frontend & Product: React, Angular, Flutter, React Native, WebRTC, Twilio

Frontend & Product: React, Angular, Flutter, React Native, WebRTC, Twilio

AI-Native Development: Claude Code, Codex, Cursor, Spec-Driven Development, Agentic Development Workflows, Parallel Agent Execution, Test/Build/Runtime Verification

Projects

Agent Orchestration Platform

- Built a full-stack AI agent orchestration platform with Next.js, TypeScript, FastAPI, PostgreSQL, Redis, and Python workers.
- Designed workflow execution, agent configuration, tool execution, Telegram webhook integration, message history, and runtime event mirroring.
- Implemented visual workflow templates, workflow runs, token/cost tracking, and OpenAI-backed agent execution with OpenClaw runtime integration.

<https://github.com/ratnadeep007/agent-orchestration-platform>

GoClaw – Agent CLI & HTTP Service

- Built an agent system supporting tool execution, LLM interaction, persistent memory, and external integrations.
- Designed CLI and HTTP service interfaces for multi-step reasoning workflows and agent runtime experimentation.

<https://github.com/ratnadeep007/goclaw>

Vidura – Self-hosted Agentic RAG System

Feb 2024 – Oct 2024

- Built a production-grade agentic RAG system with retrieval, memory, and tool execution for contextual AI responses.
- Designed multi-step agent workflows with semantic search, document grounding, and self-hosted deployment support.

<https://vidurabot.com/>

NL-to-SQL

- Built an NL-to-SQL system that converts natural-language analytics questions into safe SQL over structured datasets.
- Developed a Streamlit chat UI, Starlette/Uvicorn API, CLI, runtime schema discovery, SQL validation guardrails, and SQLite-backed demo database.
- Implemented OpenAI-based intent classification, schema-grounded SQL generation, SQL parsing/validation with sqlglot, caching, repair attempts, and automated tests.

<https://github.com/ratnadeep007/nl-to-sql>

SkyCUA – Computer Use Agent

- Built a local computer-use agent capable of interacting with system tools through containerized Docker environments.
- Designed execution workflows for task automation, tool orchestration, and controlled local agent operations.

<https://github.com/ratnadeep007/skycua>

Education

Siliguri Institute of Technology Bachelor of Technology • 7.79

Computer Science and Engineering Aug 2013 - Jun 2018