DINESH KUMAR G

Bangalore, India | +91 9148629035 | dineshkumartech3@gmail.com
linkedin.com/in/dinesh-kumar-g-5741132a4 | github.com/dkdranzer | dinesh-kumar-g-portfolio.vercel.app

SUMMARY

Software Engineer at HealthFoundry.ai building event-driven, multi-agent AI systems for healthcare. Ship distributed backend services in Python and TypeScript — LangGraph orchestration, Pub/Sub messaging, real-time delivery and human-in-the-loop routing — on GCP and AWS. Own the notification, observability and platform SDK layers used across the company's products.

TECHNICAL SKILLS

Programming Languages: Python, TypeScript, SQL, C++
AI/ML & LLM Frameworks: LangChain, LangGraph, CrewAI, LangSmith, MLflow, MCP (Model Context Protocol), Hugging Face Transformers, RAG pipelines, TensorFlow, PyTorch
LLMs & Models: Gemini, Groq, open-source models (MedASR – medical transcription)
Architecture: Event-driven architecture, microservices, distributed systems, Pub/Sub messaging, CloudEvents, Server-Sent Events (SSE), human-in-the-loop (HITL) systems, multi-agent orchestration, workflow orchestration, SDK design
Frontend & Backend: React.js, FastAPI, Node.js, Fastify, Express, REST APIs
Cloud & Infrastructure: GCP (Cloud Storage, Pub/Sub, Firebase, Firestore, Cloud SQL), AWS (EC2), Docker, Kubernetes, GitHub Actions (CI/CD)
Observability: OpenTelemetry, Grafana, Tempo, Loki, Prometheus
Databases: Firestore, Cloud SQL, MongoDB, MySQL, ChromaDB, FAISS
Healthcare & Standards: FHIR (Medplum), HL7, SNOMED CT, PHI/PII de-identification, clinical document processing
Tools: Git, GitHub, Jira, Claude Code, Cursor

WORK EXPERIENCE

HealthFoundry.ai (SynapseNet), Bangalore — *Software Engineer* **Jul 2026 – Present**

- **Designed a LangGraph-based orchestrator** that accepts work over REST and Pub/Sub and executes declarative workflows stored in Firestore — resolving the step sequence per request type (e.g. call audio), dispatching to the appropriate agent, and routing extracted results to downstream systems.
- **Developed document- and speech-intelligence agents** on Gemini 3.7 Flash — extracting structured leads, tickets and follow-ups from call recordings; clinical insights and follow-up actions from discharge summaries; and medications and conditions from handwritten and digital prescriptions.
- **Closed the loop from agent to system of record** — extracted entities are written back to the CRM, and the returned record ID is published as a notification carrying a direct deep link to the created lead, removing manual lookup from the user's workflow.
- **Built and own the observability stack** for Python (FastAPI) and TypeScript (Fastify) services using OpenTelemetry — Tempo, Loki, Prometheus and Grafana for tracing, logging, metrics and dashboards, and MLflow for LLM/AI tracing; deployed on AWS EC2 with GitHub Actions CI/CD and integrated across 8 services.
- **Authored the shared Platform SDK** — auth code exchange and Firestore profile, permission and domain lookups behind a single interface, so every service standardizes on it instead of re-implementing authentication and entitlement logic, reducing integration from bespoke wiring to a few lines.

HealthFoundry.ai (SynapseNet), Bangalore — *Data Engineer Intern* **Sep 2025 – Jun 2026**

- **Developed the Chrome extension assistant** — Firestore-based authentication resolving user permissions, profile and domain configuration — delivering actionable notifications with one-click actions: automated form filling, opening the linked record or page, and authorizing or dismissing the notification.
- **Built the Human Interaction Layer (HIL)** — an event-driven routing service that consumes CloudEvents-formatted messages from Google Pub/Sub, resolves the target recipient against Firestore-backed role, permission and domain rules, and pushes notifications in real time over Server-Sent Events (SSE), decoupling every producing agent from delivery and authorization.
- **Built the clinical document ingestion pipeline** — a LangGraph backend that pulls discharge summaries and clinical PDFs from GCS buckets, classifies them, and extracts structured data: PHI/PII de-identification with Gemini 2.5 Flash, JSON mapped to Medplum FHIR resources, and SNOMED CT diagnosis coding via HL7 terminology servers; results publish to Google Pub/Sub on a Twilio-style action schema (collect, inform, authorize, escalate).
- **Managed cloud infrastructure**: Firebase hosting for frontend deployments, Firestore and Cloud SQL as application datastores with schema design and seeding, and GCS bucket operations.

PROJECTS

LangGraph-Powered AI Chatbot **Aug 2025 – Sep 2025**

- Built a stateful, multi-threaded AI chatbot with persistent conversation memory using LangGraph and LangSmith; integrated Groq LLM with tools for web search, stock retrieval and arithmetic, plus real-time streaming, thread-based history and SQLite checkpointing via a Streamlit UI.

YouTube Video Transcript Q&A RAG **May 2025 – Jun 2025**

- Built a RAG pipeline over video transcripts — FastAPI ingestion, Cohere embeddings, FAISS vector store, ChatGroq generation via LangChain, and a Streamlit UI; Dockerized and deployed on AWS EC2.

EDUCATION

B.E. in Computer Science & Engineering — *BMS College of Engineering* **2022 – 2026 | CGPA: 8.94**

CERTIFICATIONS

Python for Data Science (NPTEL) · Machine Learning Professional Certification (RapidMiner)