

Dhruvin Rupesh Soni

AI Engineer / Applied-AI Builder | Agentic Systems + Multi-LLM | Cloud-Native Production Depth (GCP)

dhruvinrsoni@gmail.com • linkedin.com/in/dhruvinrsoni • github.com/dhruvinrsoni • dhruvinrsoni.github.io • Pune, India | Open to relocation

SUMMARY

Applied-AI engineer with 6+ years shipping production cloud-native systems, now specializing in turning AI/automation into measurable leverage. Rare trifecta: deep GCP microservices and SRE depth, a shipped and rigorously-tested multi-LLM open-source portfolio, and a documented agentic-engineering track record (custom Copilot agents, MCP, Text2SQL groundwork). Signature strength: turning hard constraints into outsized outcomes – enterprise-grade AI at \$0 incremental cost.

FOCUS

Focus: Member-of-Technical-Staff / Forward-Deployed / Applied-AI roles – agentic workflows, multi-LLM tooling, MCP, eval/scoring harnesses, and production AI delivery. Pursuing OMSCS (ML specialization).

EXPERIENCE

Senior Software Engineer — Zebra Technologies

Nov 2022 – Present

Pune, India (Hybrid)

- Engineered custom GitHub Copilot agents with quality standards codified as Markdown SOPs that autonomously generated 33,586 lines of test code in 7 days – ~80x faster than manual – with zero production incidents.
- Architected an AI “Scrum-Master” automation suite (sprint-health dashboards, auto-retrospectives, semantic ticket-deduplication), delivering enterprise-grade AI at \$0 incremental cost by ethically reusing existing authenticated LLM sessions.
- Built Model-Context-Protocol (MCP) integration over an enterprise DB2 schema, creating secure groundwork for natural-language (Text2SQL) analytics; authored a “Comparative Analysis of GenAI Strategies” that became the team’s model-selection and prompt-engineering standard.
- Led an AI-augmented test-coverage program across an Angular + Spring Boot service pair: lifted ~10-15% → 70-85% coverage (100% on a critical module), 900-1,135+ test cases over 50+ modules, at \$0 added budget.
- Re-architected a monolithic batch “god-class” into decoupled microservices and migrated legacy VM cron-jobs to Google Cloud Run under release pressure – zero critical regressions across three releases; cut review/documentation cycle time ~10x via git-context automation adopted as team SOP.

Software Engineer II — Zebra Technologies

Sep 2020 – Nov 2022

Pune, India

- Built Jenkins-on-GCP CI/CD with GKE + Helm and self-service pipelines; led SVN→Git and MongoDB→Atlas migrations off legacy infra while cutting cloud OpEx via automated off-peak pod down-scaling.
- Delivered first-in-company Google Secret Manager integration (atomic rotation, least-privilege IAM) and Trivy vuln remediation without breaking builds; built a Python MongoDB utility cutting manual effort ~95%.

Software Developer — Reflexis Systems (acquired by Zebra Technologies)

Jul 2019 – Sep 2020

Pune, India

- Built backend for an AI performance-management product (Java, Spring Boot, MongoDB, REST, Spring State Machine), shipping production features managing thousands of records.
- Drove rollout quality across a 111-client deployment: resolved 100+ issues at a 45% defect-burndown rate, lifting client satisfaction.

PROJECTS

SmrutiCortex *TypeScript, Chrome MV3, Ollama – Chrome Web Store*

Privacy-first browser history-search extension with optional local AI inference; backed by a rigorous eval harness – 1,233 unit + 45 E2E tests and 9 scorers. Demonstrates shipped, test-driven applied-AI engineering.

Samvada Studio *React 18, TypeScript, Vite – Product Hunt*

Multi-provider LLM chat UI spanning 6 providers and 35+ features with in-memory key security. Shows fluency in abstracting across frontier-model APIs and provider-agnostic LLM tooling.

Agentskills Garden *Agentic AI / skills framework*

88 hierarchical AI-agent skills demonstrating structured agentic design at scale; part of 10+ public repos plus an active engineering blog.

TECHNICAL SKILLS

Applied AI / GenAI: Agentic workflows, custom GitHub Copilot agents, Model-Context-Protocol (MCP), Text2SQL, prompt

engineering, LLM model selection, eval/scoring harnesses, RAG groundwork

LLM Tooling & Providers: Multi-provider LLM integration (6 providers), Vertex AI, OpenAI-style APIs, Ollama (local inference), in-memory key security, semantic deduplication

Cloud & Infra: Google Cloud (Cloud Run, GKE, Helm, Secret Manager, BigQuery, Cloud SQL), zero-downtime rolling updates, readiness probes, DR, IAM least-privilege

Backend & Architecture: Java, Spring Boot, REST, microservices (orchestrator pattern), Spring State Machine, DB2, MongoDB/Atlas, Python

DevOps / SRE / DevSecOps: Jenkins CI/CD, GKE + Helm, SLA-based alerting, self-healing pipelines, Trivy, secret rotation, SVN→Git & Mongo→Atlas migrations

Frontend & OSS Stack: TypeScript, React 18, Vite, Angular, Chrome MV3 extensions, Jest/E2E testing

Languages: English (professional), Gujarati (native), Hindi

EDUCATION

B.E. (Hons.), Computer Science 2015 – 2019
BITS Pilani, K. K. Birla Goa Campus
Coursework: Data Mining, Parallel Computing, Number Theory, Graphs & Networks, Computer Architecture. First Rank in Mathematics (school batch).

CERTIFICATIONS & RECOGNITION

Certifications: Google Cloud – AI Agents Edition (2025); Google Cloud – OnBoard Edition (2026); Vertex AI Prompt Design; Derive Insights from BigQuery; GC Fundamentals: Core Infrastructure; Migrate MySQL to Cloud SQL; Microsoft – VS Code Agent Mode (2025)

Awards: 5-Year Service Award, Zebra Technologies (2024); Google Cloud Skills Boost – Gold League (22,395 pts) •

Languages: English (professional), Gujarati (native), Hindi