

Dhruvin Rupesh Soni

AI / GenAI Application Engineer | LLM Agents, MCP & Prompt Engineering | Production Backend + Google Cloud
dhruvinrsoni@gmail.com • linkedin.com/in/dhruvinrsoni • github.com/dhruvinrsoni • dhruvinrsoni.github.io • Pune, India | Open to relocation

SUMMARY

AI/GenAI application engineer with 6+ years building production systems and a track record of shipping agentic LLM tooling that compounds team output. Designed Copilot agents that autonomously wrote 33,586 lines of test code in 7 days (~80x faster, zero production incidents) and an AI “Scrum-Master” automation suite delivered at \$0 incremental cost. Combines applied LLM systems (MCP, Text2SQL, prompt engineering, model benchmarking) with a hardened backend + Google Cloud foundation, plus shipped open-source LLM products.

FOCUS

Applied LLM systems, agentic workflows, MCP / Text2SQL, prompt engineering, retrieval-grounded automation

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” benchmarking leading LLMs; it became the team’s model-selection and prompt-engineering standard, and cut critical-path cycle time ~15% via an AI-first workflow.
- Led an AI-augmented test-coverage program across an Angular + Spring Boot service pair: ~10-15% → 70-85% coverage (100% on a critical module), 900-1,135+ test cases across 50+ modules, \$0 added budget.

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 infrastructure.
- Delivered first-in-company Google Secret Manager integration (atomic rotation, least-privilege IAM) and a Python MongoDB query utility that cut ~95% of manual effort.

Software Developer — Reflexis Systems (acquired by Zebra Technologies)

Jul 2019 – Sep 2020

Pune, India

- Built the 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

Samvada Studio *React 18 + TypeScript + Vite; Product Hunt*

Multi-provider LLM chat UI supporting 6 providers and 35+ features with in-memory key security for safe credential handling across model backends.

SmrutiCortex *TypeScript; Chrome Web Store*

Privacy-first Chrome MV3 history-search extension with optional local AI (Ollama); 1,233 unit + 45 E2E tests and 9 scorers for retrieval-quality evaluation.

Agentskills Garden *AI agent skills*

Library of 88 hierarchical AI-agent skills; part of 10+ public repos and an active engineering blog on applied GenAI.

TECHNICAL SKILLS

GenAI / LLM: Agentic workflows, GitHub Copilot custom agents, Model-Context-Protocol (MCP), Text2SQL, prompt engineering, RAG/retrieval grounding, LLM benchmarking & model selection, Ollama (local AI)

AI Platforms & Tooling: Google Vertex AI, Vertex AI Prompt Design, BigQuery, VS Code Agent Mode, multi-provider LLM APIs (6 providers)

Languages: Java, Python, TypeScript, JavaScript, SQL, Shell

Backend & Frameworks: Spring Boot, REST APIs, Spring State Machine, Angular, React 18, Node, Vite, microservices (orchestrator pattern)

Cloud & DevOps: Google Cloud Run, GKE, Helm, Jenkins CI/CD, Docker/Alpine, MongoDB Atlas, DB2, Secret Manager, Trivy

Practices: Agentic test generation, prompt SOPs as Markdown, zero-downtime rolling updates, SLA alerting, DevSecOps, technical documentation

Languages (Spoken): 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