Jeneeshkumar Rameshbhai Surani

AI Engineer | AI Agents, RAG & MCP-Sicherheit



jeneeshsurani@gmail.com
+49 163 7232177
Plauen, Deutschland (Umzugsbereit)
linkedin.com/in/jeneesh-surani-ai
github.com/Jeneesh1014
ai-portfolio-6u12.vercel.app

PROFIL

M.Sc.-Student der Künstlichen Intelligenz und Robotik mit **zwei Jahren Erfahrung in der Softwareentwicklung**. Ich baue **AI Agents mit LangGraph** und sichere sie mit Validierung, Guardrails und automatisierter Evaluation in der CI ab. Autor des PyPI-Pakets **mcp-guardeval**.

KENNTNISSE

AI & Agenten
LangGraph, **RAG**, **MCP**, LangChain, Pydantic, ChromaDB, Ollama, Groq API, PyTorch

Evaluation & Sicherheit
Ragas, Langfuse, OpenTelemetry, Pytest, SAFE-MCP, Red-Teaming

Entwicklung
Python, TypeScript, JavaScript, Java, FastAPI, Node.js, Spring Boot, React Native

Daten & Cloud
PostgreSQL, MongoDB, SQLite, **AWS** (EC2, S3, RDS), Docker, GitHub Actions, CI/CD

Sprachen
Deutsch B1 | **Englisch** C1

PROJEKTE

07.2026 – 10.2026

Safe MCP Agent & mcp-guardeval
Python, LangGraph, MCP, Ollama, Pydantic, OpenTelemetry, Pytest, Docker
- Lokaler LangGraph-Agent mit MCP-Zugriff auf Kundendatenbank, Wiki und Slack, getestet mit **7 SAFE-MCP-Angriffstechniken**.
- **Guardrail-Middleware mit 5 Schichten**: **alle 7 Angriffe blockiert**, unabhängig von der Modellleistung.
- Evaluations-Harness als Paket **mcp-guardeval auf PyPI veröffentlicht** (OpenTelemetry-Auswertung, Pytest-Plugin).
- Benchmark mit 48 Durchläufen: llama3.2 (3B) **100 %** der Aufgaben, llama3.2:1b nur **50 %**.

04.2026 – 05.2026

Intelligent Research Agent
Python, LangGraph, FastAPI, ChromaDB, Groq API, Tavily, Langfuse, Docker
- Agent, der pro Frage zwischen lokalen Dokumenten (RAG), Websuche oder beidem wählt: **100 % Routing-Genauigkeit**.
- Langfuse-Tracing und **41 gemockte Pytest-Tests**, die in der CI ohne API-Kosten laufen.

03.2026 – 04.2026

Ask My Docs – RAG-System
Python, FastAPI, ChromaDB, BM25, Cohere Rerank, Ragas, GitHub Actions
- Hybride Suche (BM25 und Vektorsuche) mit Reranking und **Pydantic-erzwungenen Quellenangaben**.
- **Ragas-Quality-Gate** in GitHub Actions: Unter **0,70 Faithfulness** wird das Deployment gestoppt (aktuell **0,73**).

07.2026 – 07.2026

Agentic AML Explainability Pipeline

Python, LangGraph, Ollama, Pydantic, Qwen 2.5 7B

- Pipeline mit drei Agenten (Investigator, Explainer, Auditor), die BaFin-ähnliche Berichte erstellt und bei Fehlern automatisch bis zu dreimal wiederholt.
- 4-Bit-Qwen 2.5 7B: **100 % Schema-Konformität** (20/20 Testfälle), **0,00 € Cloud-Kosten**.

BERUFSERFAHRUNG

04.2025 – 07.2025
Remote

Softwareentwickler

Pathnovo

- Mitarbeit an **4–5 Projekten in vier Monaten**: React-Native-Apps, REST-APIs, Echtzeit-Funktionen mit Socket.IO.
- Anbindung von Drittanbieter-APIs und **Deployment-Automatisierung auf AWS** mit GitHub Actions.

07.2023 – 04.2025
Indien

Softwareentwickler (Praktikum)

CodeLeap InfoTech

- **Vier Kundenprojekte als alleiniger Entwickler** von Figma-Design bis AWS-Deployment.
- Stacks: React Native, Node.js, Spring Boot, PostgreSQL, MongoDB. Optimierung von Datenbankabfragen.

AUSBILDUNG

03.2026 – 09.2027
Deutschland

M.Sc. Künstliche Intelligenz & Robotik

Hochschule Hof

Relevante Module: Applied Deep Learning, Large Language Models, Generative AI, Intelligent Robotics

10.2021 – 05.2025
Indien

B.E. Computer Engineering

Gujarat Technological University (GTU)

- **Abschlussnote: 9,04 / 10 CGPA** *(Deutsche Note: 1,3)*