

GURLEEN SINGH BHATIA

AI & Data Science Engineer | Generative AI & LLM Developer

India (Remote Ready) | +91 7070926757 | gurleensingh1608@gmail.com | [LinkedIn](#) | [GitHub](#) | [Google Dev Profile](#) | [Google Skills Profile](#)

PROFESSIONAL SUMMARY

AI & Data Science Engineer specializing in Agentic Workflows, LangGraph Orchestration, RAG Architectures, and Model Context Protocol (MCP) integration. Published researcher on sub-10B parameter autonomous agent architectures. Experienced in developing production-ready LLM application backends using Python, FastAPI, Redis, PostgreSQL, and Streamlit. Proven track record in building local-first privacy-focused AI systems, multi-agent support pipelines with human-in-the-loop review, and multi-modal vision applications with active cloud deployments. Skilled in uv, pydantic, and Docker containerization.

PUBLICATIONS & RESEARCH

Unlocking the Potential of Small-to-Medium LLMs in Autonomous Agent Architectures: A Comparative Analysis of Gemini, Llama, and Qwen
GRENZE International Journal of Engineering and Technology (GIJET), Vol. 12 (2026), Issue 2, pp. 5666–5672 | [\[Paper Abstract & PDF\]](#)

- Core Research: Evaluated sub-10B parameter LLMs (Llama 3.1 8B, Qwen 2.5 7B, Gemini 3.1 Flash Lite) operating within the Model Context Protocol (MCP) framework for enterprise edge execution.
- Complete-MCP Benchmark: Co-developed 'Complete-MCP' benchmark suite testing models on tool calling, multi-step tasks, and multi-hop reasoning graphs.
- Optimization & Key Findings: Designed adaptive parsing & rate-limiting algorithms to achieve >80% task success rates with zero data egress on local edge infrastructure.
- Enterprise Impact: Established that lightweight open-weights LLMs paired with MCP tool execution eliminate third-party API token costs while maintaining deterministic agent reliability.

TECHNICAL SKILLS

Languages & Backend: Python (AsyncIO, PyJWT, Pydantic), SQL, FastAPI, REST APIs, OAuth | **Databases:** PostgreSQL, Redis (Pub/Sub), SQLite

AI & Agent Frameworks: LangGraph, LangChain, Model Context Protocol (MCP), CrewAI, RAG Systems, Semantic Routing

LLM Platforms & Inference: Google Gemini API, Ollama / LM Studio (Local LLMs), Groq API, Google GenAI SDK

Cloud & DevOps: Docker, Docker Compose, Render, Streamlit Cloud, uv (Package Manager), Git, Linux

FEATURED AI & DATA SCIENCE PROJECTS

Multi-Agent Customer Support & Escalation System | *FastAPI, LangGraph, Redis, PostgreSQL, Docker* | [\[GitHub\]](#) | [\[Live App\]](#)

- Engineered an autonomous multi-agent support system using LangGraph stateful graph workflows (RouterAgent) for zero-shot intent categorization across Order, Billing, and Tech domains with PyJWT session state persistence.
- Built a context-rich human escalation queue powered by Redis Pub/Sub for tickets with confidence scores <0.70; fully containerized and deployed multi-service stack via Docker Compose on Render.

Smart Inbox & Calendar Automation Agent | *Python, LangGraph, Groq API, Google Calendar API, Streamlit* | [\[GitHub\]](#) | [\[Live App\]](#)

- Developed an automated email triaging agent using LangGraph to categorize incoming emails (urgent, meeting_request, newsletter) via Groq LLM inference.
- Integrated Google Calendar FreeBusy API and multi-user OAuth flow to propose conflict-free slots; engineered Human-in-the-Loop review via LangGraph interrupt() live on Streamlit Cloud.

Multimodal Vision & Content Suite (VisionAI Pro) | *Python, Streamlit, Google GenAI SDK, Gemini* | [\[GitHub\]](#) | [\[Live App\]](#)

- Created a multi-modal vision platform built with Streamlit and official Google GenAI SDK (google-genai) integrating Gemini models for structured blog posts, social copy, and e-commerce descriptions.
- Implemented interactive visual Q&A, accessibility ALT text generation (<125 chars), and multi-format session export (Markdown & TXT) live on Streamlit Cloud.

Model Context Protocol (MCP) Autonomous Data & QA Server | *Python, MCP, Scikit-Learn, NetworkX* | [\[GitHub\]](#)

- Built a local-first Model Context Protocol (MCP) server in Python exposing tools for CSV exploratory data analysis, K-Means customer segmentation (Scikit-Learn), and distribution plotting.
- Developed a Code Cartographer tool using Python AST parsing and NetworkX DAGs to map module dependencies and detect architectural coupling risks in self-healing TDD QA refactoring loops.

Local Agentic RAG System (Omni-Sentinel) | *Python, LangGraph, Ollama / LM Studio, Vector & Tabular Search* | [\[GitHub\]](#)

- Architected a zero-egress local RAG pipeline using LangGraph and a custom Semantic Router node to branch user queries between unstructured text documents and structured CSV catalogs.
- Integrated local open-weights LLMs via Ollama and LM Studio, establishing deterministic query routing and full production privacy without third-party API token costs.

EDUCATION

Master of Science in Data Science | Chandigarh University (Jul 2024 – June 2026) | **CGPA: 7.25** | Coursework: Data Structures & Algorithms, Statistics, Linear Algebra, Machine Learning

Bachelor of Computer Applications (BCA) | Arka Jain University (Aug 2021 – May 2024) | **CGPA: 8.70** | Coursework: Python Programming, Deep Learning, Neural Networks, DBMS

CERTIFICATIONS & VIRTUAL INTERNSHIPS

Virtual Internship Programs (Forage): BCG X (Data Science) • Accenture NA (Data Analytics & Visualization) • Accenture NA (Product Design) • Moreton Bay (Web Development)

Technical Certifications & Badges: Google Cloud (Generative AI Fundamentals, Responsible AI, Intro to LLMs, GenAI Explorer, Vertex AI Studio) • Kaggle (Machine Learning, Pandas) • IBM (Python for Data Science)