Aayushmaan Garg

✉ iontwoext@gmail.com • 📞 +91-9971480989 • 🔗 Portfolio • in LinkedIn • ⌨ GitHub

Education

Maharaja Surajmal Institute of Technology, GGSIPU   *Sept 2021 – June 2025*
B.Tech in Electronics and Communication Eng. (AI Specialization)
GPA: 3.6/4.0 (90.01%)

Experience

AI Engineer (Promoted from Part-Time) — Opkey, SSTS, Noida   *Jul 2023 – Aug 2025*

- Cut LLM cloud hosting costs by **92%** by redesigning AI Agents to run on 8B models instead of 235B without compromising accuracy.
- Extended Qwen3-8B model context limits by **7X** using a **Hybrid Chain-of-Agents RAG** architecture.
- Implemented **Deep Search with Qwen/Mistral** for generating web-search reports of over **180k tokens**
- Making AI **Agent Orchestration** using **Chain-of-Thought** system.
- Built and deployed RAG-based LLMs, agentic networks, and multi-agent AI with open-source (LLaMA, Mistral, Qwen) and proprietary models (GPT-3.5, GPT-4o).
- Designed LLM-driven automation generating **20K+ data rows** and training **18+ ML models**.
- Developed predictive models for current user activity from click patterns and behavior logs.
- **92%** accuracy in **PII masking** by fine-tuning BERT model for client specific data.

  Technologies & Tools Used: Python, SQL, LangChain, LlamaIndex, Hugging Face Transformers, OpenAI API, Qdrant, FAISS, Pandas, NumPy, Docker, RunPOD, Git, REST APIs, Prompt Engineering, RAG Architectures

  Skills Learned: **LLM Hosting** on a **local Data Center and Cloud**, Team Collaboration, **Training team of New Recruits** in Basics of AI

Projects

PsyFer — LLM-based Graphology Analysis (Python, Flask, Mistral 7B)   *Live*

- Compiled **50+ high-quality datasets** for handwriting psychological analysis.
- Implemented ReAct-based contextual learning system with real-time personality reports.
- Image Analysis via OpenCV and Numpy to calculate object masks and handwriting hull segments.

SmartHub — Multi-URL Doc Intelligence (FastAPI, Qdrant, BeautifulSoup, Mistral)   *Archived*

- Built SmartAgent handling **100+ URLs** concurrently with context-aware responses.
- Scraped and processed **500K+ data points** with advanced NLP pipelines.
- Integrated Google Safe Browsing API for live URL threat detection.
- Implemented Spider Crawler for mesh scraping.

Scaffold — Golang CLI Tool   *GitHub*

- Created CLI to scaffold framework-specific project structures, cutting setup time by **70%**.
- Supported templates for Flask, Gin, and other frameworks — ensuring consistent, clean codebases.

Certifications and Skills

AI/ML: LLM Integration, RAG, AI Agents, Fine-tuning, NLP Semantic Search, Anomaly Detection
Certifications: IBM Data Science Professional, IBM AI Engineering
Language Skills: Japanese JLPT N4 (B1 level)