

Aryan Gupta

✉ aryan.gupta210302@gmail.com | 💻 aryangupta | 🐙 aryangupta | 🌐 aryangupta01.github.io | 📞 +91-9981060403

Education

Indian Institute of Technology Kharagpur

Kharagpur, India

Integrated M.Sc. Economics (CGPA: 8.08/10)

2020 – 2025

Work Experience

Eightgen.ai | *AI Engineer Intern*

Dec 2024 - Feb 2025

Multimodal Financial RAG System

- Developed and deployed a **multimodal financial RAG system** using **Colpali embeddings** on **Qdrant Vector DB**.
- Optimized retrieval for **20,000+** documents, achieving **40x speed** improvement using **mean pool embeddings**.
- Deployed **inference servers** on **AWS EC2** and **payload filtering** with **Qdrant clustering**.
- Integrated **OpenAI LLM APIs** to deliver **contextual query responses** from retrieved data.

Financial AI Agents Deployment

- Developed **Financial AI agents** for a chatbot using **LangGraph** and **LangSmith**, for intelligent query resolution.
- Engineered a scalable **REST API** with **asynchronous production-grade** code, integrating **async SQLite** for **long-term memory** retention and conversation continuity.
- Built **Selenium based REST API** for **automated real-time** stock market financial data web scraping.

Walmart Global Tech | *Software Development Engineer Intern*

May 2024 - July 2024

- Integrated **Prometheus** server & **Grafana** on **Walmart Cloud Native Platform**, enhancing monitoring abilities.
- Developed **custom metrics** and **Grafana** dashboards for a **Spring Boot** application, improving system visibility.
- Created a **RESTful API** for **PUT** calls to update data in the service portal, streamlining data management.

Indian School of Business | *Machine Learning Engineer*

Feb 2024 - March 2024

- Engineered the development of a **SOTA text clustering system** pioneered for **open-ended complaints analysis**.
- Utilized **k-means** for optimal k, **HDBSCAN** for flexible clustering, and **DistilBERT-based** word **vectorizations**.
- Improve complaint categorization using **Silhouette** and **Adjusted Rand Index** scores for better clustering.

Projects

Story Maker Agent | AI-Powered Children's Storybook Generator | Multi-Agent System

June 2025

Objective: Build A.I multi-agent system for automated children's storybook creation with AI illustrations

- Architected **5-agent CrewAI framework** orchestrating **GPT-4**, & **DALL-E 3** for collaborative story generation.
- Build **RESTful FastAPI** backend with **async task** handling and **real-time status tracking** for scalable operations.
- Engineered **Pydantic-based** custom tools for LLM, enabling **DALL-E 3** image generation & **mdpdf**-based PDF export
- Deployed **API** with **Swagger** documentation, **environment configuration**, and **concurrent task management**.

Multimodal RAG System | Document Intelligence with LLaVA | Vision-Language Pipeline

May 2025

Objective: Build multimodal RAG system for document analysis with vision-language understanding

- Designed **end-to-end RAG pipeline** using **LLaVA 7B**, **LangChain**, & **ChromaDB** for unified text-image retrieval.
- Build **PDF ingestion** system with **Unstructured.io** for automated text, tables, & images extraction with **OCR**.
- Integrated **MultiVectorRetriever** storing **embeddings** for search & **raw content** generation via **InMemoryStore**.
- Built **base64 image processing pipeline** with auto resizing, format detection, & **vision-language prompt setup**.
- Deployed **Ollama** on **Colab** with **xterm** integration & configured **HuggingFace** embeddings (**BAAI/bge-large-en**)

Text-to-SQL Query Generator | LoRA Fine-tuned LLM | Natural Language Processing

June 2025

Objective: Instruction Fine-tune Qwen3-0.6B via LoRA for natural language to SQL query translation

- Engineered **text-to-SQL pipeline** using **Qwen3-0.6B** with **LoRA fine-tuning** (rank-16), on **20k samples** dataset.
- Implemented **SFTTrainer** with **custom conversation** formatting, **completion-only** training, & **FP16** optimization.
- Achieved **80% match** accuracy & **98.0 BLEU score**, boosting base performance by **75.2 points** in translation quality

Technical Skills

Programming Languages

C/C++, Python, SQL, Java, HTML, CSS

Frameworks & Databases

Tensorflow, Langsmith, Langchain, Pytorch, Hugging Face, Mysql, SQLite, Fastapi

Tools, Platforms & Libraries

Linux, Git/GitHub, Docker, AWS EC2, Pandas, Sklearn, Matplotlib, Keras, Numpy, NLTK

Positions of Responsibility

Kharagpur Data Analytics Group, IIT Kharagpur | *Senior Member*

2022 - 2023

- Organized & coordinated workshops on Machine Learning & **GAN** assisted **Neural Art** for students across country.
- Organized national Kharagpur Data Science Hackathon with 3500+ registrations & sponsored prize pool of 1.25 lac INR.