Saud Ahmad

Data Scientist & AI Engineer | Specializing in
LLMs, RAG, And Agentic AI Systems

Email: sauds6446@gmail.com
Links: GitHub LinkedIn
Contact: 📞03462936273
Address: Kabal, Swat, Pakistan

[Profile]

Data Scientist and AI Engineer focused on building intelligent, retrieval-augmented, and autonomous AI systems. Skilled in LLM development, contextual retrieval, agentic reasoning, and scalable model deployment using LangChain, Hugging Face, and modern MLOps stacks. Passionate about creating adaptive AI ecosystems that merge data science, language understanding, and real-world automation.

[Projects and Experience]

CV Evaluation & RAG Assistant System (CoterGlobal) – AI-Powered HSE Recruitment Platform (COTER Global) – Built full-stack recruitment system with AI CV evaluation, intelligent assistant, and management dashboard. Implemented FastAPI async backend, PostgreSQL with 3,500+ candidates, OpenAI GPT-4 integration, FAISS vector search, automated data normalization, duplicate detection, real-time dashboard with 11+ filters, requirements tracking, R2 storage, and responsive frontend. Deployed on Railway. Tech: FastAPI, LangChain, OpenAI API, PostgreSQL, FAISS, asyncpg, Cloudflare R2, Python AsyncIO, HTML/CSS/JavaScript. (github_link)

RAG AI Assistant – Built a retrieval-augmented generation assistant that reads and analyzes CVs and GitHub repositories to deliver recruiter-ready summaries and responses using FAISS and GPT models. (github_link)

Retail Insight Generator – Built a Streamlit-based AI app that converts natural language queries into SQL for real-time retail analytics. Leveraged LLMs for natural language–to–SQL translation, enabling non-technical users to access data insights effortlessly. Tech: Streamlit, OpenAI GPT, SQL, Python. (github_link)

Agentic AI Systems – Developed autonomous agents capable of executing real-world tasks such as web scraping, resume parsing, and conversational interaction, leveraging modern frameworks like LangChain and OpenAI APIs. (github_link)

Duplicate Question Detection (BERT) – Implemented a BERT-based NLP model to detect semantically similar questions, including data preprocessing, model training, and evaluation. (github_link)

Breast Cancer Classification – Designed a deep learning model to classify tumors as malignant or benign using the Scikit-learn breast cancer dataset. (github_link)

Diabetes Prediction – Created a predictive model using SVM to identify potential diabetes cases based on medical data.

Titanic Survival Prediction – Applied feature engineering and classification algorithms to predict passenger survival outcomes. …….. (github_link)

Traffic Accident Analysis (Pakistan) – Conducted exploratory analysis and data visualization to extract insights on accident frequency and fatalities. (github_link)

Data Warehousing Project – Designed a full ETL pipeline involving data extraction, cleaning, transformation, and loading into a MySQL data warehouse. (github_link)

Data Mining – Documented hands-on experiments in data mining, analysis, and model evaluation across diverse datasets. (github_link)

Access Vision – Created a face recognition-based entry and exit management system integrating camera input and sensor data for automated access control. (github_link)

OpenCV Face Recognition – Developed real-time face recognition using OpenCV and the face_recognition library for identity verification. (github_link)

OpenCV Basics Collection – Compiled foundational scripts demonstrating image reading, color manipulation, and contour detection. …… (github_link)

Bus Reservation System (C++) – Implemented an object-oriented bus management console app with booking and record-tracking functionalities (github_link)

Quiz Tic-Tac-Toe Game (C++) – Combined a quiz mechanism with classic tic-tac-toe to enhance interactivity and logic design (github_link)

Smart Robot Car (Arduino) – Built a multi-mode robot car integrating sensors and motor control, demonstrating embedded programming and automation. (github_link)

[Skills]

Core AI & Machine Learning

- Machine Learning: Supervised & Unsupervised Learning, Predictive Modeling, Feature Engineering
- Deep Learning: CNNs, RNNs, Transformer Architectures (BERT, GPT, LLaMA, Mistral)
- Natural Language Processing (NLP): Text Classification, NER, Summarization, Question Answering
- Large Language Models (LLMs): GPT-based Systems, Fine-tuning, Prompt Engineering, Ollama
- Retrieval-Augmented Generation (RAG): Context Retrieval, Knowledge Integration, Hybrid Search
- Agentic AI / Autonomous Agents: Multi-agent Orchestration, Tool Use, Memory Management, Reasoning Loops
- Generative AI: Text, Code, and Multimodal Generation
- Model Optimization: Parameter-Efficient Tuning (LoRA,, QLoRA).
- Embeddings & Vector Databases: Text Embeddings, Semantic Search, Vector Indexing (FAISS, Chroma, Pinecone, Milvus)

Backend & API Development

- FastAPI: Async APIs, Pydantic models, dependency injection, RESTful design
- Async Programming: Python AsyncIO, async/await patterns, concurrent task execution
- Session Management: Redis for state persistence, session serialization/deserialization
- Document Processing: PDF extraction (PyPDF2), DOCX parsing (python-docx), async file operations (aiofiles)
- Web Servers: Uvicorn, Flask, Gunicorn

Frameworks & Libraries

- LLM & Agent Frameworks: LangChain, LangGraph, LlamaIndex, Semantic Kernel, CrewAI
- Modeling & NLP: Hugging Face Transformers, OpenAI APIs, Anthropic, Ollama
- Deep Learning: PyTorch, TensorFlow, Keras
- Data Science & Analytics: scikit-learn, spaCy, Pandas, NumPy, Matplotlib, Seaborn, Plotly
- Web Development: Streamlit, HTML, CSS, JavaScript, Jinja2

MLOps, Deployment & Tools

- MLOps: Docker
- Model Serving & APIs: FastAPI, Flask, Streamlit
- Version Control & Environment Management: Git/GitHub, venv, Conda
- Cloud: AWS EC2, S3

Programming Languages

- Python (ML, AI, automation)
- SQL (data querying)
- C++
- R (statistical analysis)
- JavaScript (frontend development)

Mathematics & Theoretical Foundations

- Calculus: Differential & Integral Calculus, Partial Derivatives, Optimization of Loss Functions, Backpropagation
- Linear Algebra: Vectors, Matrices, Eigenvalues/Eigenvectors, Singular Value Decomposition (SVD), Matrix Factorization for Embeddings
- Probability & Statistics: Descriptive and Inferential Statistics, Hypothesis Testing, Regression Analysis, ANOVA, Correlation, Sampling Methods, Probability Distributions, Bayesian Inference, Confidence Intervals, and Statistical Significance Testing
- Optimization Methods: Gradient Descent, Stochastic Optimization, Regularization Techniques, Convex Optimization
- Mathematical Modeling: Analytical reasoning for predictive models and probabilistic systems

Business Intelligence (BI) & Visualization

- Power BI, Tableau, Looker, Google Data Studio
- Dashboard Design, Data Storytelling, KPI Tracking
- Integration with AI Insights & Automated Reports

[Professional & Soft Skills]

- Strong Communication & Presentation
- Team Collaboration & Leadership
- Analytical & Critical Thinking
- Time Management & Adaptability
- Continuous Learning & Research Mindset
- Problem Solving & Creativity

[Certifications]

IBM – What is Data Science?
Foundational understanding of data science concepts, tools, and workflows across ML, DL, big data, and cloud computing.

IBM – Tools for Data Science
Hands-on experience with Python, R, SQL, Jupyter, and RStudio; practiced Git/GitHub and cloud-based data science tools.

IBM – Data Science Methodology
Applied the CRISP-DM framework for project structuring, model selection, data preparation, and storytelling.

IBM – Python for Data Science, AI & Development
Built strong Python foundations with Pandas, NumPy, and Jupyter; applied to REST APIs, automation, and web scraping.

IBM – Machine Learning with Python
Implemented supervised and unsupervised algorithms using scikit-learn; developed and evaluated ML models through projects.

DeepLearning.AI – Generative AI for Everyone
Studied generative AI concepts, prompt engineering, and responsible AI practices for real-world applications.

[Education]

BSc in Data Science GIK 2025