

New York, USA | AyushChhoker15@gmail.com | +1 (680) 230-9058

[GitHub](#) | [LinkedIn](#) | [Portfolio](#)

SUMMARY

Data Scientist and Machine Learning Engineer specializing in Retrieval-Augmented Generation (RAG) systems, streaming analytics, and advanced NLP applications. Experienced in building end-to-end ML pipelines, integrating multiple LLM APIs, and developing scalable data platforms. Strong foundation in deep learning, graph neural networks, and automated document generation with proven research contributions in medical AI.

EDUCATION

Master of Science in Data Science and Analytics | August 2024 – May 2026

SUNY Polytechnic Institute, Utica, NY

Relevant Coursework: Machine Learning, Deep Learning, Graph Neural Networks (GNN), Data Visualization, Natural Language Processing, Big Data Analytics

Master of Computer Application | August 2020 – September 2022

Galgotias University, Greater Noida, India

Relevant Coursework: Software Engineering, Data Structures, Algorithms, Database Management Systems

PROFESSIONAL EXPERIENCE

Graduate Assistant – SUNY Polytechnic Institute | October 2024 – Present

Research & Development in Advanced AI Systems

- Architecting advanced Retrieval-Augmented Generation (RAG) systems to enhance AI-driven information retrieval with 40% improved response accuracy
- Building Graph RAG architectures leveraging Neo4j for multi-hop reasoning and contextual understanding in large language model applications
- Designing synthetic dataset generation pipelines for training and evaluation of machine learning models across diverse domains
- Engineering robust data pipelines using Python and cloud services for processing and integrating heterogeneous data sources into RAG frameworks
- Collaborating on NLP research involving embedding techniques, semantic search optimization, and advanced prompting strategies

Data Analyst – Stop-Not Services | August 2022 – August 2024

Market Research & Business Intelligence

- Managed end-to-end research lifecycle for 15+ market research projects using IBM SPSS Statistics, Python, and advanced statistical modeling
- Developed 10+ interactive dashboards using Power BI and Tableau, improving stakeholder decision-making efficiency by 25%
- Conducted comprehensive data services including cleaning, processing, and validation using automated scripting and manual techniques

- Identified market trends and opportunities through data-driven insights, contributing to revenue growth for 20+ clients
- Maintained 99.5% data quality and research integrity across multiple concurrent projects

TECHNICAL PROJECTS

DataFlow Intelligence Platform

Streaming Analytics & Multi-LLM Integration

- Architected end-to-end streaming analytics platform using Python, Azure PostgreSQL, and Streamlit for real-time data processing
- Integrated multiple LLM APIs (OpenRouter, GPT-4, Claude, Gemini) for intelligent response optimization and automated analytics
- Implemented machine learning components for outlier detection, comparative analysis, and predictive insights in logistics workflows
- Designed scalable ETL pipelines handling large-scale data with optimized performance and sub-second query response times
- Tech Stack: Python, Azure, PostgreSQL, Streamlit, LangChain, REST APIs

End-to-End RAG Chatbot System

Multimodal Conversational AI

- Developed production-ready chatbot with LangChain, Haystack, Ollama, ChromaDB, and Gradio interface
- Enabled multimodal document retrieval with conversational memory across sessions using vector databases
- Optimized embedding retrieval achieving sub-second response times across 50,000+ document vector database
- Implemented advanced prompt engineering techniques including Chain-of-Thought and Few-Shot Learning
- Tech Stack: Python, LangChain, Ollama, ChromaDB, Gradio, NLP, Vector Databases

TECHNICAL SKILLS

Programming & Analytics: Python • SQL • R • Dart • Statistical Modeling • Hypothesis Testing

AI & Machine Learning: TensorFlow • PyTorch • Scikit-learn • Deep Learning • Computer Vision • NLP

AI & Prompt Engineering: Large Language Models • Prompt Optimization • Chain-of-Thought Prompting • Few-Shot Learning • RAG Systems

Frameworks & Tools: LangChain • Haystack • Ollama • Streamlit • Gradio • Flask • LlamaIndex

Databases: PostgreSQL • MySQL • MongoDB • Neo4j • ChromaDB • Pinecone • Vector Databases

Data Visualization: Tableau • Power BI • Matplotlib • Seaborn • Plotly

Cloud & DevOps: Azure • AWS (S3, Lambda, EC2) • Docker • Git • GitHub Actions • CI/CD

Development Tools: VS Code • Jupyter • PyCharm • Anaconda • Android Studio • VMware

CERTIFICATIONS

- Advanced Machine Learning Specialization | Coursera
- Machine Learning with Python | IBM
- Supervised Machine Learning: Regression and Classification | DeepLearning.AI
- Google Data Analytics Professional Certificate | Coursera | February 2022