

COLTRANE [REDACTED]



EDUCATION

[REDACTED]
Bachelor of Science in Computer Science
Specialization in Data Science and Analytics

Graduation Date: August 2022

SKILLS

Languages: Python, SQL

Frameworks: Langchain, Langgraph, CrewAI, Pydantic, Pandas, Streamlit, Flask, FastAPI, Scikit-Learn, Matplotlib, Keras, PyTorch

Tools: Spark, Docker, Kubernetes, Podman

Platforms: Cursor, Visual Studio Code, Jupyter Notebook, AWS, Grafana, Prometheus, Bitbucket, Github, IBM Watsonx, Postman, GCP, Databricks

Technical Skills: Generative AI, Machine Learning, Natural Language Processing, Analytics

Soft Skills: Communication, Leadership, Design Thinking, Project Management

WORK EXPERIENCE

[REDACTED]
AI Engineer

February 2025 - Current

- **Generative AI:** Designing and scaling Generative AI applications from PoC to production across multiple content generation workflows and RAG-based systems, **boosting productivity by at least 50%**.
- **Language Model Optimization:** Reducing LLM inference costs by **at least 40% through prompt and context optimization**, while maintaining **at least 90% accuracy on business content generation metrics**.
- **Backend Development:** Building scalable backend APIs with FastAPI using object-oriented design and asynchronous event loops, achieving **at least 60% lower latency and approximately 20% cloud cost savings through improved resource utilization**.
- **Deployment and Observability:** Deploying dockerized Generative AI systems on AWS using ECR and ECS for scalable orchestration. Integrating CloudWatch, Grafana, and MongoDB for real-time monitoring and logging. Designing and implementing custom LLM observability metrics, including token usage, cost, latency, and error tracking, which **reduces on-call incidents by 50%** and **improves MTTR by 60%**, significantly lowering operational overhead and improving reliability.

[REDACTED]
AI Engineer - Client Engineering

April 2023 - February 2025

