Ayush Chaudhary

ayushch1183@gmail.com | in/ayushch1183 | github.com/Ayush-Chaudhary

Education

University of Maryland College Park, MD
Masters of Science in Applied Machine Learning **(GPA: 4.0)** Expected Graduation: May 2025

Indian Institute of Technology New Delhi, India
B.Tech in Electrical Engineering with Minor in Computer Science **(GPA 9.1/10)** May 2023

Technical Skills

Languages	**Proficient:** Python(5 yrs), C++(4 yrs), MATLAB, **Familiar:** C, Java, JavaScript, Dart
NLP & ML	LLMs, Retrieval-Augmented Generation (RAG), Embeddings, Vector Databases, PyTorch, TensorFlow, Transformers, Computer Vision, Generative AI
Tools & Tech	Pinecone, Tree-sitter, AST Parsing, Anthropic AI, Code Analysis, AWS (S3, SageMaker, Lambda, API Gateway), MLflow, Docker, Kubernetes

Industry Experience

Mirage Lab, UMD College Park, MD
Machine Learning Researcher June 2024 - Present

- Developed an **AI-driven metrology pipeline** integrating **Line Segment Detector**, **ElDet CNN**, and a **fine-tuned LLM (GPT-4)** to automate dimension analysis on grayscale images.
- Optimized the **model training pipeline** to generalize across diverse CAD designs, leveraging **distributed training** and **CUDA-accelerated inference**, achieving **95%** time reduction and improving uncertainty from **120 μm to 40 μm**.
- Enhanced system efficiency with **parallelism strategies** and **MLFlow**-tracked experiments, enabling rapid **hyperparameter tuning** and **model optimization** for real-time industrial applications.

Mastercard Gurgaon, India
Machine Learning Engineer June 2022 - July 2022

- Designed and optimized a **scalable data processing pipeline** for **real-time fraud detection**, leveraging **parallel computing** and **vectorized operations** to process **100k+** data points, boosting report generation speed by **80%**.
- Fine-tuned **Copynet** and **GPT-2** for **financial anomaly detection**, integrating **transformer-based feature extraction** to enhance **money laundering risk assessment** accuracy by **30%**.
- Developed an **automated ML deployment pipeline** using **Docker**, **Kubernetes**, and **CI/CD workflows**, eliminating manual intervention and saving **1800+** analyst hours annually.

Research Experience

Advanced Code Intelligence System with RAG College Park, MD
Independent Research Project Jan 2025 - Mar 2025

- Engineered a **Retrieval-Augmented Generation (RAG) system** leveraging **Anthropic AI embeddings** to enable natural language querying of GitHub codebases, achieving context-aware understanding through **parse tree analysis**.
- Developed a sophisticated **code-chunking algorithm** using **Tree-sitter AST parsing** to maintain semantic integrity of code snippets, optimizing for context retention with **token-aware segmentation**.
- Implemented a scalable **vector search solution** with **Pinecone** for high-dimensional semantic code retrieval, with **namespace-based organization** for efficient multi-repository management.

AI-Powered Product Recommendation System on AWS College Park, MD
Research Assistant with Dr. Samet Ayhan Aug 2024 - Dec 2024

- Engineered an **adaptive recommendation pipeline** leveraging **AWS SageMaker**, integrating **Neural Collaborative Filtering (NCF)** and **XGBoost** to improve ranking precision.
- Optimized model deployment using **AWS Lambda, API Gateway, and Step Functions**, reducing inference latency by **30%** and enabling **real-time A/B testing** with **SageMaker Endpoint Variants**.
- Implemented **model drift detection** via **Amazon CloudWatch, S3 event triggers, and SageMaker Model Monitor**, ensuring continuous retraining and adaptation to evolving user preferences.