

ASHUTOSH CHOUDHARI

4ashutosh98@gmail.com • (412) 626-9743 • [LinkedIn](#) • [GitHub](#) • [Portfolio](#) • Pittsburgh, PA

EDUCATION

CARNEGIE MELLON UNIVERSITY, Pittsburgh, PA

Aug 2022 - Dec 2024

Master of Information Systems Management

CGPA: 3.78/4.00

Courses: Computational Data Science, Distributed Systems, Unstructured Data Analytics, Time Series Forecasting, Machine Learning

VELLORE INSTITUTE OF TECHNOLOGY, Vellore, India

Jul 2016 - Jun 2020

Bachelor of Technology - Electronics and Instrumentation Eng (Minor in Computer Science and Eng)

CGPA: 3.94/4.00

Courses: Database Management systems, Data Structures, Neural networks, Python programming, Statistics, Operating Systems

PROFESSIONAL EXPERIENCE

AI/ML Engineer, Pyramyd (Remote)

Jun 2025 - Present

- Designed a **transformer-based ML model in PyTorch** to predict ratings for software product reviews achieving 87% accuracy.
- Developed **AI-powered RFP Requirements Extractor** using **o4-mini LLM** and **Azure AI Search**, implementing parallel processing architecture to extract, categorize requirements from proposal documents with **90%+** accuracy across documents.
- Optimized performance for 5x concurrent processing with vector search for intelligent document cross-referencing.

Software Developer, Confirmed LLC (Pittsburgh, PA)

Jan 2025 – Present

- Developed microservice using **LangChain, FastAPI, Azure Container Apps** for nametag scanning, achieving **92% accuracy**.
- Integrated Python service with .NET backend via **REST APIs**, optimized SQL query performance by **43%** through **NL-to-SQL**.
- Built **CI/CD pipeline** in **Azure DevOps** with pytest testing and **Docker** deployment; reduced deployment time by **70%**.

Machine Learning Engineer (Student Consultant), Nexteer Automotive (Remote)

Aug 2024 – Dec 2024

- Created document comparison application to extract, compare, and analyze PDF documents using **Azure AI Document Intelligence, Adobe PDF Extract API** and **OpenAI's GPT-4o**, enhancing document management efficiency and saving **2 FTEs**.
- Implemented user-friendly Streamlit GUI with a robust backend data handling system leveraging MongoDB for persistent caching, to minimize API call costs and improve processing speed by 5 times.

Data Scientist Intern, MedSarathi LLC, (Remote)

Jun 2023 - Aug 2023

- Developed an A/B tested, regression-weighted recommendation engine, increasing residency acceptance by **14%**.
- Built a full-stack web app (**GPT-4, React, Flask, PostgreSQL, Heroku**) for streamlined service delivery and advanced analytics.

Solution Delivery Analyst, Deloitte (Hyderabad, India)

Jun 2021 - Jul 2022

- Developed a **Python-based IOC triage tool** to process 500+ IPs, domains, URLs, and emails daily, classifying inputs, and integrating OSINT APIs (VirusTotal, WHOIS, AbuseIPDB), reducing average query time by 65% with parallel execution.
- Architected a Flask REST backend with PostgreSQL and Redis**, exposing endpoints for IOC normalization, rate-limit handling, and API key management, enabling **multi-user scalability and cutting redundant OSINT calls by 40%**.
- Built a React-based frontend with secure API integration** to provide dashboards, results tables, and export workflows (Markdown/PDF/CSV), ensuring **consistent and auditable threat intelligence reporting across environments**.

PROJECTS

Network Security Phishing Detection System (*MongoDB, MLflow, AWS ECR, EC2, Docker, FastAPI, CI/CD*)

Aug 2024 – Nov 2025

- Developed a machine learning pipeline for phishing detection using 7 tuned classification algorithms and achieving 96% F1-score and 97% precision with the best-performing model (**XGBoost**), representing a 15% improvement over baseline models.
- Architected scalable MLOps infrastructure with **FastAPI** microservices, GitHub Actions **CI/CD**, **AWS ECR containerization**, and **MLflow** experiment tracking, reducing deployment cycles by 78% while maintaining 99.2% system availability.

Artificial Intelligence Policy Chatbot (*Python, Hugging Face, PyTorch, Large Language Models, Generative AI*)

Jan 2024 – Mar 2024

- Fine-tuned **Llama-2 7B** on a 1000 document AI-policy corpus using **PEFT** and **LoRA**, built a preprocessing/tokenization pipeline, and ran hyperparameter sweeps to achieve 23% BLEU, 18% ROUGE, and 21% METEOR gains over baseline.
- Engineered a LangChain powered RAG pipeline integrating a ChromaDB vector store and dynamic prompt templates to surface top-ranked policy excerpts, boosting answer relevance by 30% and cutting average inference latency by 25%.

Question-Answering System Development (*Azure Machine Learning, PyTorch, nltk, spaCy, SentenceTransformers*)

Sept 2023 – Nov 2023

- Fine-tuned a pre-trained **BERT model** for question-answering tasks, achieving 68% accuracy on the SQuAD test set, a **28%** improvement over the pre-trained model.
- Deployed the fine-tuned BERT model as a scalable cloud-based question-answering service on **Azure Container Instances (ACI)**, leveraging Azure ML for deployment, achieving efficient inference on unseen data.

SKILLS

Programming languages: Python, Java, SQL, R, JavaScript

Frameworks / Libraries: Numpy, Pandas, Scikit-learn, PyTorch, Tensorflow, NLTK, Hugging Face, Flask, LangGraph, LangChain

Tools: Tableau, Jira, GitHub, Docker, MongoDB, Neo4j, FAISS, Chroma, Ollama, Sqlite, Nvidia NIM, Pinecone

Certifications: AWS Certified Cloud Practitioner, Cert. of Cloud Security Knowledge, Agile Product Owner, Agile Scrum Master

Cloud tools: AWS Bedrock, SageMaker, Lambda, S3, API Gateway, Azure ML studio, Vertex AI, GCP