

ASHUTOSH CHOUDHARI

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

Software/AI/ML Engineer and Data Scientist with 3+ years of experience designing automation pipelines, LLM driven solutions, and cloud native applications. Strong engineering background with expertise in microservices, and distributed processing. Adept in NLP, data science, deep learning, and MLOps, delivering reliable, production ready solutions that drive measurable business impact.

EDUCATION

CARNEGIE MELLON UNIVERSITY, Pittsburgh, PA

Aug 2022 - Dec 2024

Master of Information Systems Management

CGPA: 3.78/4.00

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

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)

Feb 2025 - Sep 2025

- 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 Science Intern, MedSarthi 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

Auto Emails for Referrals (Python, Google Workspace APIs; Gmail, Sheets, Drive, LLMs, GitHub Actions)

Apr 2025 - Jun 2025

- Automated Gmail + Google Sheets + Google Drive referral pipeline, trimming manual outreach time per email by **~80%**.
- Delivered 1000+ personalized sends with zero duplicates via sheet-driven dedupe, resume mapping, and role-aware templates.
- Orchestrated multi-provider LLM drafting (GitHub Models, OpenAI, Azure) with dynamic prompts, fallbacks for consistent tone.
- Built 4-hourly CI/CD with token health checks, alerting, and dry-run safety, sustaining **98%+** workflow uptime.

Search Engine LLM (LangChain, Groq, HuggingFace, Streamlit, Ollama, GitHub Actions)

Jan 2025 - Feb 2025

- Developed an AI agent-powered search chatbot using Llama3 LLM with LangChain, Groq, and HuggingFace, integrating Arxiv, Wikipedia, and DuckDuckGo search tools to fetch information, handling multi-hop queries with up to 3 external sources.
- Deployed the application on Hugging Face Spaces with CI/CD automation via GitHub Actions, reducing deployment time by **40%**. Implemented context retention for up to 5 user interactions, improving conversational coherence and response accuracy.

Network Security Phishing Detection System (MongoDB, MLflow, AWS ECR, EC2, Docker, FastAPI)

Aug 2024 - Nov 2024

- 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.

SKILLS

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

Frameworks / Libraries: PyTorch, Tensorflow, NLTK, Hugging Face, scikit-learn, NumPy, pandas, XGBoost, flask, LangChain

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

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 AI Foundry, GCP Vertex AI, BigQuery