Azam Khan

New York, NY, 10025
+1 (646) 647-9552 | ak4973@columbia.edu | https://github.com/azamkhan99

EDUCATION

Columbia University New York, NY
M.S. Computer Science, Machine Learning Track, **GPA: 3.90** Expected May 2025
- Relevant Courses: Cloud Computing, NLP, Practical Deep Learning Systems

University of Edinburgh Edinburgh, GB
B.Sc. Artificial Intelligence and Computer Science, **First Class Honors** Jun 2021
- Relevant Courses: Data Mining & Exploration, Data Structures & Algorithms, Applied ML

WORK EXPERIENCE

Columbia Climate School/NASA GISS New York, NY
ML Researcher Sep 2024 - Present
- Train accurate and interpretable neural networks and probabilistic models to predict water quality, utilizing Google Earth Engine and TensorFlow for development, contributing to climate impact assessment in the NYC Watershed.

NeoStats Analytics Dubai, AE
Consultant - Data Engineer Feb 2024 - May 2024
- Utilized Azure Data Factory and Databricks to migrate 88 SAP tables and ingest external API data into Azure Synapse, achieving a 5x faster dashboard refresh time while building a Data Warehouse for a **consumer company**.

Deloitte UK London, GB
Data Engineer Sep 2021 - Nov 2023
- Spearheaded development of a CO2 emissions calculator for a **global technology company**, adopting Streamlit for prototyping. Mentored a junior consultant in Docker, Git, and GCP to implement a CI/CD pipeline.
- Developed backend for end-to-end LLM-powered RAG-based document chatbot, leveraging PaLM2 model, LangChain, and ChromaDB. Enabled accurate information extraction from complex regulatory documents.
- Designed and implemented cross-functional data architecture and executive dashboards for public policy teams across 8 EU member firms, developing scalable ETL pipelines with Airflow, Python, and AWS (S3, Lambda, Athena) for a **global e-commerce leader**, resulting in informed decision-making.
- Created APIs and ETL pipelines using Airflow, Flask, and MySQL to integrate data from third-party REST APIs into a **global e-commerce leader's** risk-profiling application, enabling near real-time monitoring of tactical risks.
- Advised a **major UK-based telco** on cloud storage and infrastructure usage optimization, identifying areas of inefficiency and proposed recommendations, driving potential cost savings of up to 350k annually.
- Deployed the Deloitte TrueVoice platform for a **large UK retail bank** and an **international hospitality chain**, building data pipelines on AWS with PostgreSQL databases. Trained NLP models achieving 88% accuracy, improving call-handling quality and customer experience.

SKILLS

Programming & Data Science: Python, SQL, Spark, Pandas, TensorFlow, Sklearn, NLTK, LangChain, Pytorch
Data Infrastructure & Cloud: MySQL, Postgres, MongoDB, DynamoDB, FastAPI, Airflow, AWS, Azure, Databricks
DevOps & Visualization: Docker, Jenkins, Git, Jira, Tableau, Streamlit
Certifications: Microsoft Certified: Azure Data Engineer Associate

PROJECTS AND RESEARCH

DESDR Research Team (Professor Dan Osgood and Professor Lydia Chilton) Jan 2025 - Present
- Collaborated on an open-source toolkit at Columbia Climate School, developing APIs, databases, and infrastructure for training accurate disaster risk models using Flask, Postgres, and Python.

MeetSynthia.ai Jan 2025 - Present
- Developed and tested guardrails to eliminate prompt engineering for brand-specific LLM chatbots, optimizing performance, role-based instructions, and governance for more effective AI-engineered prompts.

Internet Real Time Lab (Professor Henning G. Schulzrinne) Sep 2024 - Jan 2025
- Built LLM agent using RAG and DSPy to protect data and privacy online, alerting users of potential vulnerabilities. Created a browser extension in JavaScript and Flask to extract and process user-agreements.