Alan Mohan

alanmohan@cmu.edu • linkedin.com/in/alan-mohan8

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
M.S Artificial Intelligence Engineering-Information Security (GPA: 3.91) — August 2025 – December 2026 (Expected)

Birla Institute of Technology and Science, Pilani — Dubai, UAE
B.E Computer Science (GPA : 9.43 / 10) — September 2021 - July 2025

TECHNICAL SKILLS

Programming: Python (primary), Java, JavaScript, PHP, C/C++, SQL
AI/ML & GenAI: Generative AI, Large Language Models (LLM), Prompt Engineering, Explainable AI (XAI), PyTorch, Keras, LangChain, Vector Databases, Retrieval-Augmented Generation (RAG), HuggingFace, MCP, LLM Interpretability, LLM Evals
Cloud Infrastructure: AWS, GCP, Docker, CI/CD, Git, Linux, Flask

EXPERIENCE

Bank of New York — Pittsburgh, PA
AI INTERN — May 2026 –August 2026

- Built and deployed a scalable AI-agent evaluation platform to automate evals across the AI Hub, reducing time spent on manual evaluation.
- Built an end-to-end workflow for evals with features like dataset management, synthetic test case generation, online evaluation and agent performance analysis.
- Developed an extensible LLM evaluation engine with 30+ metrics including RAGAS, LLM-as-a-Judge, and multi-turn agent metrics to measure response quality, goal completion, knowledge retention, and tool-use accuracy.

Carnegie Mellon University — Pittsburgh, PA
GRADUATE RESEARCH ASSISTANT — January 2026 – May 2026

- Designed a mechanistic interpretability pipeline using PyTorch and HuggingFace Transformers to analyze how LLMs encode decision-making features for tasks such as comparing mutual funds, addressing the black box problem.
- Developed linear probing experiments to quantify information encoding in transformer activations, achieving >90% probe accuracy and identifying where model reasoning breaks down.
- Conducted layer-wise activation analysis to map financial metrics across layers, contributing to ML interpretability research by revealing specific encoding peaks.

Condo Protego — Dubai, UAE
AI INTERN — February 2025 - July 2025

- Designed and implemented AI-driven workflow automation solutions, streamlining critical business processes
- Built and deployed a Python-Flask + React app with Generative AI integration on GCP Cloud Run, cutting sales proposal document creation time by 70%
- Developed automation agents for data extraction and PDF processing, saving 12 hours per employee weekly

DigiProctor — Remote
GENERATIVE AI AND BACKEND DEVELOPER INTERN — June 2023 - August 2023

- Built Medical Referral Prioritization System to triage medical referrals using GPT-3.5 for analysis & AWS Textract for data extraction in Python, reducing document triaging time from 10 mins to ~30s
- Applied prompt engineering to optimize GPT-3.5 model performance
- Developed web app in PHP CodeIgniter & jQuery, ensuring smooth user experience
- Collaborated with CEO, PM, and senior developers from design to deployment

PROJECTS

Explainable AI based Deep Learning Models for Intrusion Detection of IoT

- Developed Intrusion Detection Systems with CNN, LSTM models, achieving up to 99% accuracy on 4 datasets
- Applied explainable AI techniques such as SHAP and LIME for explainability and feature selection
- Published a research paper in an Elsevier Q1 journal (DOI: 10.1016/j.compeleceng.2025.110256)

Detecting Sentiment Steering Attacks on RAG-enabled LLMs

- Built framework to detect/mitigate sentiment-steering attacks via data poisoning in RAG systems
- Achieved 97.75% recall in detecting poisoned passages on WikiASP dataset
- Proposed scalable defenses, securing RAG pipelines against manipulation

- Published in IEEE ICC-2026 conference (DOI: 10.1109/ICC59461.2026.11588193

AWARDS

- 3rd Rank RIFT: Capture the Flag by Google Developer Groups, 2025
- UAE Winner and Global Nominees NASA International Space Apps Challenge 2023
- Rank-9 in BITS Pilani Postman API Hackathon 2.0

ADDITIONAL SKILLS

Problem-Solving & Analytical Thinking, Leadership & Team Collaboration, Project Management, Research & Technical Writing