Muhammad Abdullah Bilal

FAST-NUCES | abdullah-bilal.com | +923477054002

EXPERIENCE

EKAI.AI (MIT C10 LABS) | Data Scientist (Founding Engineer)
Nov 2024 – Present | Remote, US

- Built a service that migrates legacy data warehouse code to dbt. Used LangGraph to coordinate multiple agents with shared state, and Neo4j to track lineage so the agents had the right context when rewriting each file. According to Ekai design partners, it led their warehouse migration efforts.
- Built a service that links entities and metrics across dbt, Snowflake, Oracle, etc. semantic models. Matches on names, glossary terms, keys, lineage, and value patterns, scores each match by confidence, and explains why it made the call.
- Built a code dependency tracker in Neo4j that flags downstream files affected by a change, cutting down on regressions during edits. Reduced the instances where downstream fixes were required by almost 70 percent.
- Built a GraphRAG pipeline to reduce the context window size over transformation code so the team could reuse logic buried in existing SQL and Python files. Wrote custom LangChain runnables for the retrieval and reranking steps. Reduced the number of redundant questions asked by the user across model generation by 30 percent.
- Developed system that generates dbt code, utilizing the DBML schemas to create ETL pipelines to reduce the dev time. Has a system (using neo4j) to track DBML changes so other workflows could pick up the latest state. The entire Ekai product is built around this workflow.

RAYN GROUP | Data Scientist (L2)
June 2023 – Nov 2024 | Islamabad, PK

- Developed a financial and login fraud detection system capable of handling up to 4 million transactions per day with an F1-score of 94% for one of the largest financial institutions in Pakistan. Used pytorch, with ELK stack for data ingestion.
- Developed a Geospatial tracking and analysis system capable of tracking and generating detailed performance metric reports for up to 90,000 devices simultaneously in real time. Used Google Maps API and implemented the solution in C++ to reduce processing time by 85 percent.
- Founded and led Rayn's AWS DeepRacer team, maintaining the top position in Pakistan for five consecutive months. Used Reinforcement learning for training AWS DeepRacer car.

PROTECT LAB | Software Developer
May 2022 – June 2023 | Islamabad, PK

- Developed a Cyber-Threat correlation detection system for a Threat Detection System for the Government of Pakistan.

KDD LAB | Research Intern
September 2021 – May 2023 | Islamabad, PK

- Worked on graph compression algorithm comparisons and worked on graph compresion with bounded error.

RESEARCH

FAST | Research Paper (Currently under review)
Collaborated with **Jawad Jaral** on building a methodology for fixing security vulnerabilities in codebases by breaking them down to an abstract syntax tree (AST) and using RAG to find the top-k candidates from a dataset of known CVEs and using specialized coding LLMs to fix it.

KDD LAB | Researcher
Jan 2023 – May 2023 | Islamabad, PK
Worked with **Kifayat Khan (PhD)** and **Sungyoung Lee (PhD)** on a paper **PNRG: Knowledge Graph-Driven Methodology for Personalized Nutritional Recommendation Generation**, which presents a tool that provides personalized recommendations for CKD patients.

INTRODUCTION

Data Scientist with experience in machine learning, data analysis, and AI-driven solutions across cybersecurity, health-tech, fintech, and Generative AI. Proficient in Python, SQL, NLP, deep learning, and statistical modeling. Strong analytical, problem-solving, and communication skills, with a passion for continuous learning and leveraging data-driven insights.

EDUCATION

FAST-NUCES

Bachelor of Science in Computer Science
Aug 2019 - June 2023 | Islamabad, PK
• CGPA: 3.38 / 4.0
• First student in FAST-NUCES to complete the program in 7 semesters
• Completed program with extra credits

SKILLS

PROGRAMMING

4+ years
• Python • C/C++ • Java

TECHNOLOGY

LLM Frameworks
•LangGraph •LangChain •RAG •Ollama •pgvector •Embeddings

Data & Databases
•SQL •dbt •Postgres •Neo4j •Cypher •ETL •DBML

Data Science / ML
• PyTorch •scikit-learn •Pandas •Polars •NumPy •NLP

Geospatial
•OSM •Google Maps API

Cloud & Infra
•AWS •Docker •Git

CERTIFICATIONS

•Neo4j Graph Data Science Certification
•Neo4j Certified Professional

LINKS

Github: **github.com/abdullahbilal64**
LinkedIn: **linkedin.com/in/muhammad-abdullah-bilal**
Email: **hello@abdullah-bilal.com**