

ABDULLAH ATHAR

Machine Learning Engineer

+44-7501-411469

ama86@cantab.ac.uk

[LinkedIn](#)

Cambridge, United Kingdom

SUMMARY

AI/ML Engineer with ~3 years at GlaxoSmithKline, specializing in ML solutions. Successfully delivered impactful projects, including a recommender system for genetic targets, RNA structure prediction pipelines, and compound-target activity prediction models. Proficient in multi-view ML and unsupervised Bayesian frameworks. Published in multiomics data integration. Integrated Masters in Engineering from the University of Cambridge, ranking 10/371. Proactive problem solver with a passion for cutting-edge ML techniques.

EDUCATION

University of Cambridge

Masters in Information and Computer Engineering
2017-2021

- Starred Distinction (10th in a cohort of 370)
- Prize for best ML project

Lahore Grammar School

2014-2016

5 A* and Distinguished Cambridge Learner Award

SKILLS

Programming Languages: Python, R, SQL, Bash

SWE: Git, TDD, Unix, SLURM, ETL, Docker, Jira, Confluence, CI/CD, FastAPI, React, Node.js

ML: Scikit-Learn, PyTorch, PyG, Dask, MLflow, Lightning, Vertex AI

Cloud: AWS, GCP, Azure

Databases: MySQL, PostgreSQL, MongoDB

PROJECTS

Contributed to the closed-source software **Stradview** to develop statistical shape models of the human pelvis and answer the research question: *Does hip joint angle affect 3D joint space width?*

Phenonaut: multiomics data integration for phenotypic space exploration (Shave et al., 2023).

Cambridge Engineering ML project: Used model predictive control to perform a swing-up of a cartpole system around its unstable equilibrium. Won a prize for best ML project in my year group.

PROFESSIONAL EXPERIENCE

AI/ML Engineer II

GSK | 2023 - Present

- Developed a recommender system with LambdaMART, trained on genetic and disease features to rank genetic targets for specific disease areas, allowing research teams to accelerate the selection process by 25%.
- Developed scalable ETL pipelines for RNA structure prediction datasets leveraging Dask to enable processing at TB scale.
- Developed a Transformer-CNN architecture with ALiBi positional encoding for predicting RNA accessibility, surpassing thermodynamic models by 25% on MAE.
- Published in multiomics data integration, contributing to cutting-edge research in the field.

Data Science and Machine Learning Associate

GSK | 2021 - 2023

- Deployed a FastAPI micro-service for GSK Global Health, leveraging PostgreSQL to query demographic information and delivering predictions on health inequalities using a pre-trained model.
- Developed image-omics contrastive learning methods using TripletLoss to learn joint embeddings from high-content datasets, resulting in a 50% increase in benchmark metrics.
- Designed relational data schemas for DNA encoded technology as one of the initial data scientists on the team and streamlined data access by reducing query execution time by 30%.

Software Engineering Intern

BAE Systems UK | Jun-Sep 2019

Striker II helmet mounted display team:

- Enhanced performance of product simulators by 30% through profiling tools and algorithmic optimisation.
- Applied SOLID software development principles to create reusable and scalable code with classes representing distinct concepts akin to components in the display.
- Developed a foundational understanding of the hardware components behind AI/ML technologies