

Muhammad Usman

Bruchwiesenanlage 4, 66125, Saarbruecken, Germany | +49 174 276 7793 |
usmanchaudhary.eng@gmail.com | LinkedIn | GitHub



Professional Summary

AI Engineer for Bioinformatics specializing in **RAG-powered knowledge systems**, **AI agents for multi-omics and NGS**, and **cloud-based ML platforms**. Bioinformatician and Machine Learning Engineer with experience building scalable, reproducible pipelines across **genomics, transcriptomics, and proteomics**. Currently in the final phase of a Master's degree, seeking opportunities at the intersection of **biological data analysis, machine learning, and infrastructure**, with a focus on translating complex biological data into actionable insight.

Education

Bioinformatics, M.Sc. Oct 2022 – Present

Saarland University, Germany

- Thesis: *ML-Driven Exploration of NRPS Engineering in E. coli*
- Focus: Computational Biology, Genomics, NGS, Machine Learning

Biomedical Engineering, B.Sc.

Oct 2018 – Aug 2022

University of Engineering and Technology (UET), Lahore, Pakistan

GPA: 3.72/4.00 with Honors

- Thesis: *ECG Analysis: Monitoring of Signs & Symptoms*

Work Experience

AI Safety Research Fellow

Nov 2025 – Present

Saarland University, Germany

- Developing **multi-omics ML models** for **personalized oncology** using fusion and graph-based approaches.
- Assessing **uncertainty, causality, and robustness** under cohort and distribution shifts.
- Building reproducible **Snakemake + Docker pipelines** for **AI-safety benchmarking**.

Machine Learning & Bioinformatics Engineer

Jun 2024 – Present

Myria Biosciences, Switzerland

- Processed **10,000+ genomic and proteomic samples**, reducing pipeline runtime by **35%** on **HPC systems**.
- Benchmarked **variant calling pipelines**, improving **SNP sensitivity by 8%**.
- Designed and automated **ETL, NGS, and ML pipelines** (Python, SQL, Docker, CI/CD), ingesting **50 GB/week** with **<2% error**.
- Built a **graph-based knowledge system (Neo4j)** integrating **10,000+ entities**, improving query performance by **50%**.
- Delivered **30+ dashboards and analytical reports**, supporting decisions across **3 clinical programs**.

Research Assistant

Oct 2024 – Jul 2025

Saarland University, Germany

- Applied **AI-driven CADD methods** for **biosynthetic pathway optimization** and molecule screening.
- Integrated **synthetic biology** and **computational chemistry** workflows for in-silico experimentation.
- Automated **ML pipelines** using **Python, PyTorch, and scikit-learn**, improving reproducibility.

Data Scientist in Drug Bioinformatics

Jun 2024 – Dec 2024

Helmholtz Institute (HIPS), Germany

- Integrated **NGS and LC-MS/MS proteomics data**, enabling discovery of **15 novel BGCs**.
- Developed **ML models** improving prediction accuracy by **22%**.
- Applied **graph-based methods** with **experimentally validated predictions**.
- Built reproducible **Snakemake pipelines**, saving **40+ hours/month**.

Research Assistant

Dec 2022 – Sep 2024

Korean Institute of Science & Technology (KIST) Europe, Germany

- Supported **sequencing-driven synthetic biology projects**, validating **100+ cloning experiments**.
- Standardized workflows and trained students, reducing onboarding time by **30%**.
- Performed **ELISA, cloning, and phage-host interaction** studies for bacteriophage engineering.

Volunteering

Research Mentor, Computational Biology

Sep 2022 – Present

UET Lahore, Pakistan

- Mentored undergraduate projects applying **machine learning and deep learning** to **biological data** (EEG, genomic, physiological).
- Guided students in **research design, feature engineering, and model evaluation**.
- Provided hands-on mentorship in **Python, scikit-learn, and PyTorch**, emphasizing **reproducible coding practices**.

Projects & Certifications

- **LangChain Mastery: Agentic AI Projects, Agentic Workflows** and LLM orchestration
- **RNA-seq Multi-Omics Workflow, End-to-end NGS pipeline**
- **Containerized Single-Cell Analysis, Reproducible scRNA-seq workflows**
- **FastAPI CI/CD Reference, Automated Docker deployment**
- **Heart Disease Prediction and Analysis, 92% accuracy** (RF / XGBoost)
- **NeuroVision Explorer, MRI analysis using CNNs**
- **Neo4j Certified Professional, Graph Data Science Certified**
- **AWS Cloud Practitioner Essentials**

Publications

Reprogramming Filamentous FD Viruses to Capture Copper Ions. <i>ChemBioChem</i> . DOI: 10.1002/cbic.202400237	2024
Bacteriophage Engineering for Improved Copper Ion Binding. <i>Macromolecular Bioscience</i> . DOI: 10.1002/mabi.202300354	2023
Electrocardiogram Data Visualization and Dimensionality Reduction. <i>ICBHS 2022</i> .	2022
Brain Tumor Segmentation and Detection on MRI Images. <i>ICBHS 2022</i> .	2022

Awards & Hackathons

Winner, Health Hack Saar Oct 2025
Saarbrücken, Germany

MEDPlex: Bridging the Gender Data Gap in Pediatrics

- Built a real-time **evidence-based pediatric medication recommendation system** addressing **gender bias in dosing**.
- Delivered a **full working prototype within 48 hours** with potential to improve medication safety and reduce dosing errors.

Finalist, ODDO BHF Equity Research Hackathon

Nov 2025

Paris, France

Enhancing Equity Research Client Engagement

- Selected among **top teams** and invited to pitch at **ODDO BHF Headquarters**.
- Built an **AI-augmented platform** improving analyst-client engagement via availability signalling and automated follow-ups.

Participant, Defensive Acceleration Hackathon

Nov 2025

Global

BioCast AI: Global Disease Outbreak Forecasting

- Developed a hybrid **LSTM + XGBoost** epidemiological forecasting pipeline with **PCA/clustering analytics** and outbreak risk prediction.

Skills

NGS & Bioinformatics: WGS, WES, RNA-seq (bulk & single-cell), methylome, variant calling, annotation, Bioconductor

ML & Data Science: Scikit-learn, PyTorch, PCA/UMAP/t-SNE, clustering, Random Forest, SVM, XGBoost, SHAP, GNNExplainer, graph ML

DevOps & Infrastructure: Docker, Git, Bash, Linux Server Management, CI/CD, Workflow Automation, Cloud Deployment

Pipelines & Tools: Snakemake, Nextflow, BioPython, Pandas, NumPy, Jupyter, Flask

Data Engineering & Databases: SQL, ETL/ELT, API pipelines, Neo4j knowledge graphs, data modeling

Visualization: Plotly, R Shiny, Jupyter Dashboards, RMarkdown, interactive reporting

Languages: English (C2), German (A1), Urdu (C2), Punjabi (C1)