WASIM AFTAB, PhD

Bioinformatician & Software Engineer

+491724685100
wasim.aftab@med.uni-muenchen.de
Neuried, Germany
LinkedIn GitHub ResearchGate

EXPERIENCE

Post-doctoral researcher, Ludwig Maximilian University of Munich January 2022 - Present

- **Functional Genomics and NGS Data Analysis:** Programmed Snakemake pipelines for bulk, single-cell RNA-seq, and ATAC-seq data using R and Python which reduced the data processing time by **50%** and enhanced the reproducibility of results.
- **GWAS-based Genomics:** Familiar with PLINK, GWAS catalog, Genotype Imputation server for GWAS analysis.
- **Knowledge Management using LLM:** Developed a biomedical QA web app using Python, JS, HTML/CSS that improved information retrieval by **137.5%** over the state-of-the-art method and deployed it on a DigitalOcean droplet.
- **Prompt Engineering:** Applied prompt engineering on GPT-4 to extract entities/relationships for knowledge graph construction.

Research Assistant during PhD, Ludwig Maximilian University of Munich September 2015 - December 2021

- **Biomedical Data Integration:** Developed algorithm and software to integrate Imaging Mass Spectrometry (IMS) with shotgun proteomics data for peptide identification in IMS.
- **Proteomics Data Analysis:** Analyzed mass spectrometry data for protein complex prediction, peptide identification, and gene expression network visualization.
- **R packages and Desktop App Development:** Built R packages, Shiny, and Electron apps used by **10+** labs.

Senior Software Engineer, Icelero Technologies Private Limited April 2012 - December 2012

- **Test Automation Development:** Built software testing pipelines with Java, Selenium, and AutoIt, which significantly accelerated testing processes and reduced manual effort.

Scientific Programmer, Jawaharlal Nehru Centre for Advanced Scientific Research August 2009 - March 2012

- **HPC Optimization:** Optimized C subroutines, boosting computational speed for quantum chemistry simulations.
- **Protein Function Prediction:** Developed a C program to predict the functions of hypothetical proteins using structural features from the Protein Data Bank.

Research Intern, Jawaharlal Nehru Centre for Advanced Scientific Research February 2009 - June 2009

- **C program to Diagonalize Large Sparse Matrices:** During the internship, I worked on my final year Bachelor's degree thesis, which involved developing a C program to diagonalize large sparse matrices, a crucial tool for solving quantum chemistry problems, particularly those involving large-scale configuration interaction calculations of electronic wave functions.

SUMMARY

- **Collaborations:** Participated and drove interdisciplinary, collaborative projects, resulting in multiple publications in top-tier science journals.
- **Problem-solving:** Solved complex problems across diverse domains
- **Web & Desktop App Development:** Developed end-to-end software implementations by integrating backend and front-end components.
- **Biomedical Data Science:** Expertise in genomics, proteomics data analysis, and biomedical data integration. Developed R packages, Shiny, and Electron apps used by 10+ labs.
- **Generative AI Application Development:** Developed and deployed web apps using GenAI technologies to facilitate biomedical information retrieval. Applied prompt engineering on LLMs to construct knowledge graphs, enhancing biomedical information retrieval and domain-specific QA.
- **Application Deployment:** Deployed web applications on DigitalOcean droplets and local lab machines; experimenting with AWS and Kubernetes for cloud computing.

SKILLS

- **Machine Learning & Data Science:** Supervised & Unsupervised Learning (Random Forest, XGBoost, SVM, K-means), Deep Learning (CNN, RNN, GANs), Bayesian Data Analysis, Predictive Modeling, NLP (Transformers, BERT, GPT), Model Deployment and Optimization, Mathematics, Mixed Models, Method Development.
- **LLM Technologies:** OpenAI API, Ollama, Retrieval-Augmented Generation (RAG), Graph-RAG, Knowledge Graph Integration with LLM, Prompt Engineering, Fine-Tuning.
- **Programming & Scripting Languages:** R, Python 3, MATLAB, C, JavaScript, HTML/CSS, Shell scripting, SQL.
- **Frameworks & Libraries:** PyTorch, Hugging Face Transformers, Scikit-Learn, NumPy, Pandas, BeautifulSoup, R Bioconductor, Linux, Ubuntu, RStudio.
- **Data Visualization & Deployment:** Plotly, Matplotlib, ggplot2, Cytoscape, R Shiny, AWS (EC2, S3), DigitalOcean, Docker, Git.

LANGUAGES

- **English:** C1 (Advanced)
- **German:** B1 (Learning)

CERTIFICATIONS

Genomic Data Science Specialization (Coursera), Quantitative Proteomics (EMBL), Challenges in Multiomics Data Integration (EMBL), Cypher Fundamentals (Neo4j), Building Knowledge Graphs with LLMs (Neo4j), Mathematics for Machine Learning: Linear Algebra (Coursera), Mathematics for Machine Learning: Multivariate Calculus (Coursera).

EDUCATION

PhD (Specialization: Bioinformatics), Ludwig Maximilian University of Munich, Munich, Germany — September 2015 - December 2021

MSc Computer Engineering (Specialization: Machine Learning), King AbdulAziz University, Jeddah, Saudi Arabia — January 2013 - May 2015

BE Computer Engineering (Specialization: High Performance Computing), Visvesvaraya Technological University, Belgaum, India — September 2005 - August 2009

PORTFOLIO

https://wasimaftab.de