

Sheharyar Akhtar

Lahore, Pakistan | +923015388946 | sheharyar7146@gmail.com | [linkedin.com/in/sheharyarakhtar/](https://www.linkedin.com/in/sheharyarakhtar/)

Skills

- **Databases:** MySQL, Snowflake, GreenPlum, PostgreSQL, MongoDB, Athena
- **Programming Languages:** Python, R Programming, C++, MATLAB, Stan, SQL, noSQL
- **Data Analysis & Machine Learning:** Python (Scikit-learn, PyMC, Matplotlib, Time series), R, Flask/ Dash, LLMs (LangChain, LangGraph, CrewAI, OpenAI, Gemini, GPT, Llama)
- **Cloud:** Azure (Machine Learning Studio, Azure OpenAI, Prompt Flow, AI Search), AWS (Glue, Athena, S3, EC2, SageMaker, Lambda, AI Services)
- **Tools:** RStudio, Jupyter Notebook, Anaconda, Git, Docker, MS Excel, MLOps, MLFlow
- **Techniques:** EDA, Causal Inference, Regression, Classification, Clustering, A/B testing, Forecasting, Orchestration
- **Soft Skills:** Agile, Scrum, Leadership, Communication, Problem Solving, Strategic Planning

Experience

Turing

Sept '24 – Present

Sr. Data Scientist/ Team Lead

- Leading a team of 7 LLM trainers and mentoring them on shepherding the LLM's thought process from issue understanding to patch implementation.
- Collaborate with Meta on enhancing the Llama 3.2 model by guiding the LLM through debugging scenarios using their novel AI Debugger platform.
- Provide insights and data that contribute to the training of the next version of the Llama model, improving its reasoning and problem-solving capabilities.
- Work closely with AI researchers to align training processes with model development goals, while pioneering innovative LLM training methods.

Systems Limited

Aug '23 – Nov'24

Consultant – Data Science

- Streamlined Azure DevOps pipeline with Docker for efficient development and deployment, collaborating cross-functionally for standards adherence and user-centric enhancements.
- Developed various PoCs surrounding a plethora of data science domains including credit scoring, churn modeling, RAG and multi-agent workflows
- Deeply assessed clients' needs based on their vision presented in RFPs to architect solutions, taking budget constraints in mind, on various cloud platforms including Cloudera, Azure, AWS and GCP

Afiniti

Apr '21 – Aug'23

Data Scientist II - Applied AI

- Trained and led up to 5 resources across different accounts across the Iberia region to define strategic roadmaps and ensure high-quality deliverables and timely work.
- Transformed raw data feeds into meaningful modeling datasets using stored procedures in PostgreSQL and MySQL
- Developed interactive RMarkdown/ Shiny reports and Apache Superset dashboards to track and visualize important KPIs
- Engineered customer features that improved model validations by over 50% when included in the modelling process

Projects

Diversity Forecasting Initiative

- Led a Diversity Forecasting Initiative to address gender bias in management, targeting a 45% representation of female Full-Time Employees (FTEs) through predictive models utilizing tree-based algorithms.
- Engineered interactive simulations allowing exploration of counterfactual scenarios, empowering decision-makers to visualize and assess the impact of strategic changes on female FTE percentages in management.

Technologies: Python, lightgbm, scikit-learn, Plotly-dash, Azure (ML Studio, AppService, DevOps), MLFlow, Git, Spark

LLM Consultant using Multi-Agent Workflow

- Created a RAG architecture using Azure AI Search to index and vectorize document text and images for retrieval augmented generation
 - Created a multi-agent workflow for assessing RFP requirements, researching relevant literature online and recommending state-of-the-art solutions that satisfy the clients' needs
- Technologies:** CrewAI, Azure OpenAI, Azure AI Search, Azure Machine Learning, LangChain, Google Serper API, custom tools

Education

B.S. Mechanical Engineering – GIKI

- **Courses:** Engineering Statistics, Numerical Analysis, Intro to Computing, OOP, Linear Programming, Software Engineering

Achievements and Awards

- Published a Peer-reviewed article in the Cureus Journal of Medical Science, contributing to understanding the root cause of hesitation in getting the Covid Vaccine – Cited 27 times
- Authored multiple articles for the online teaching blog "Towards Data Science," sharing insights on Bayesian modeling – Over 10k views