Fatima Farooq

Arlington, 76013 | +1 817-970-3202 | fatimafarooq2310@gmail.com

Fatima Farooq | LinkedIn

Objective

Seasoned Senior AI Developer and Data Scientist with a Master's in Data Science and a Bachelor's in Software Engineering. Currently a PhD student and Graduate Research Assistant in the Large-Scale Intelligent Data Systems (LIDS) Laboratory at UTA, conducting cutting-edge research on knowledge graphs and graph neural networks. Offering over five years of experience in machine learning, NLP, computer vision, and deep learning. Eager to leverage expertise in Python, PyTorch, Keras, and geospatial analysis in a dynamic remote work environment. Aiming to contribute to innovative projects with a global impact.

Professional Experience

Graduate Research Assistant

Large-Scale Intelligent Data Systems (**LIDS**) Laboratory, University of Texas at Arlington, USA | Aug 2024 – Present

- Lead research on scalability and optimization of Graph Neural Network (GNN) architectures applied to large-scale, heterogeneous knowledge graphs, emphasizing efficient link prediction and node classification tasks.
- Investigate scalability challenges of heterogeneous graph neural networks (e.g., RGCNs) on large-scale knowledge graphs and propose optimization strategies.
- Develop and benchmark novel graph sampling and partitioning strategies to enhance training efficiency and scalability on real-world heterogeneous knowledge graphs.

Senior AI Developer

Center of AI and Computing (**CENTAIC**), NASTP, Pakistan | Jan 2022 – July 2024

- Successfully spearheaded a project on the classification of aircraft trajectories using air defense radar data, developing and fine-tuning a robust model that effectively handles real-time, unseen time series data for critical predictive accuracy in air defense applications.
- Led the development and implementation of advanced anomaly detection models to identify unusual patterns in defense radar data, significantly enhancing threat detection capabilities.
- Utilized PyQGIS to manipulate and analyze geospatial data, improving the accuracy of geospatial anomaly detection algorithms.
- Developed and maintained software applications using Qt, ensuring high usability and performance across multiple platforms.
- Implementing Graph Databases on noisy geospatial data using Memgraph and Cypher to retrieve and manipulate data.
- Engineered AI-driven decision support tools for air defense applications, demonstrating substantial improvements in decision accuracy and response times.
- Implemented architectural elements and Inter Process Communication (IPC) with stress performance testing.

Lab Instructor

FAST - National University of Computer and Emerging Sciences, Isb, Pak | Jan 2019 - Jan 2022

- Facilitated hands-on learning in Data Analysis, Visualization, and Database Systems, significantly enhancing student comprehension and practical skills.
- Provided expert instruction in Object-Oriented Programming and Digital Logic Design, receiving

commendation for teaching effectiveness.

Software Engineer - Deep Learning

University of Engineering and Technology, Taxila | Apr 2018 – July 2019

- Collaborated on the development of a driving simulator for autonomous vehicles, leading to innovative data collection methods.
- Implemented and optimized machine learning algorithms for object detection and lane detection, contributing to significant advancements in virtual simulation technologies.
- Designed and developed a Driving Simulator using OpenCV to collect data for model training.
- Contributed to projects focused on Urban Planning, Driving Simulator, and Road Safety.
- Organized workshops and seminars on Virtual Reality, including a seminar at APES'18, UET Lahore with 100+ participants.

Education

PhD Computer Science

University of Texas at Arlington, United States | Aug 2024 – Present

- Conducting research on knowledge graphs and their application using Graph Neural Networks (**GNNs**), with a focus on improving **link prediction** and **node classification** tasks.

- Investigating the scalability and computational complexity of various GNN architectures, such as **RGCNs**, to optimize performance on large-scale, multi-relational datasets.

MS Data Science

FAST - National University of Computing and Emerging Sciences, Pakistan | Sept 2020 – Aug 2022

- Graduated with a **Silver Medal**, specializing in Deep Learning, Computer Vision, and Natural Language Processing.
- **Thesis**: *"Transformer-based Automatic Speech Recognition for Low Resource Languages"* achieved notable recognition for innovation.

CGPA: 3.74/4.00 **(Silver Medalist).**

BS Software Engineering
University of Engineering and Technology, Taxila, Pakistan | Oct 2015 – Aug 2019

- Focused on Deep Learning, Computer Vision, and Autonomous Vehicles.

Thesis: *"Autonomous Driving Simulator using OpenCV and Deep Learning."* achieved **1st prize** in Open House and Job Fair 2019

CGPA: 3.48/4.00.

Key Skills

- **Technical:** Expert in Python, R; proficient in PyTorch, TensorFlow, Keras; experienced with Apache Hadoop, Spark, Pig, Hive; Advanced skills in PyQGIS and Qt for developing sophisticated GIS applications and user interfaces.
- **Domains:** Machine Learning, AI, Computer Vision, NLP, Data Analysis, Graph Databases, Anomaly Detection. Specialized in leveraging PyQGIS for geospatial data analysis and anomaly detection in diverse datasets.
- **Tools:** Proficient in Git, QGIS, Hugging Face, OpenCV. Skilled in using Qt for creating cross- platform

software applications. Experienced with PyCharm and Visual Studio Code (VSCode) for efficient code development and project management.

- **Soft Skills:** Exceptional at remote collaboration, problem-solving, and time management.

Achievements & Distinctions

- **GRAND PRIX AWARD** (winner), The 6th CPWC, Tokyo Japan 2018.
- **Silver Medal**, MS Data Science, FAST – NUCES, 2022.
- **First Prize:** Awarded for the top Final Year Project at U.E.T. Taxila's Open House & Job Fair 2019.

Projects

- **Decision Support Tools for Air Defense:** A machine learning-based framework to enhance decision-making in air defense systems.
- **ASR for Low resource Language (Urdu):** A transformer-based NLP based model to write transcription of language with 16% WER only.
- **Driving Simulator for Autonomous Vehicles:** Utilized Deep Learning and OpenCV for simulating autonomous driving scenarios.
- **Big Data insights:** Text Mining on Unstructured Pokec Dataset that includes Data preprocessing, Association rules, Frequent Pattern mining and LSH Algorithm.
- **Movies recommendation system:** Using Collaborative Filtering that handled 25 million datasets and used content-based recommendation on input features.
- **Future Prediction and analysis:** on DBLP Dataset and predict journals and conference papers and Data visualization through t-SNE. Achieve 93% accuracy using models and test using neural networks.
- **Data analysis using Hadoop Pig:** Most visited pages in different categories from Visits Dataset
- **Image Restoration System"** using MATLAB to restore and enhance images when distorted with imperfect lens.
- **Automating Loan Eligibility Prediction:** Using Data Analysis and Machine Learning, preprocessing, model building using techniques like Logistic Regression, Decision Trees, and Random Forests
- **Analysis of Geographic Spread:** and Impact of COVID-19 Pandemic - A global perspective in the First Six Months and Dashboard for visualization.
- **Comprehensive Dashboard:** for Monitoring and Analyzing COVID-19 Spread and Impact Worldwide - Using D3 – A Data Driven Document Library.
- **University Database Management System:** encompassing student records, course details, tutor information, and various relationships between them, facilitating efficient data organization and retrieval.
- **Dashboard on Movie recommendation system**: using D3 library with HTML, CSS and JavaScript. Revenue, Budget and Popularity of the movies from year 2009 to 2019.
- **Darknet2020**: Analysis and Visual insights from Big Data, Analyzing network traffic, safe inferences about the network in the Darknet 2020 dataset.
- **Heart Disease Prediction**: using Machine Learning Techniques - Four supervised approaches namely Naïve Bayes, Logistic Regression, Decision Tree and Random Forest
- **Part time Projects:** Have done 200+ Projects related to Deep Learning, Applied Data Science, Machine Learning, Image Processing, Visualization on Fiverr in last 5 years.