Fazli Imam

AI Researcher (Level II) at MBZUAI
Website Google Scholar Github LinkedIn E-mail
(+971) 54 317 5850

Education

Mohamed Bin Zayed University of Artificial Intelligence Aug 2022 – May 2024
Master of Science in Machine Learning (Fully funded scholarship) *GPA 3.60/4.00*

Sri Lankan Institute of Information Technology (SLIIT) Jan 2016 – Dec 2020
Bachelor of Science (Hons) in Information Technology with specialization in Data Science *GPA 3.81/4.00*

Technical Skills

Programming Languages & Libraries: Python, R, SQL, Scikit-learn, XGBoost, BeautifulSoup, Selenium, LangChain, FastAPI, Streamlit, Flask NumPy, Pandas, PyTorch, Keras, TensorFlow
AI/ML Frameworks & Tools: HuggingFace, OpenAI API, MLflow, RAGs, Weights & Biases, Google Analytics
Data Engineering & Platforms: ETL, Databricks, Docker
Cloud Platforms: AWS (S3, EC2, SageMaker), Microsoft Azure (ML Studio, Databricks), Google Cloud Platform
Data Visualization Tools: Tableau, Power BI, Seaborn, Matplotlib
Developer Tools: Git, Bash scripting

Experience

Machine Learning Researcher Jul 2024 – Present
Mohamed Bin Zayed University of Artificial Intelligence

- Led research initiatives on multimodality and visual-temporal reasoning in vision-language models notably experimenting with **LLAMA**, **LLAVA**, **GPT-4o**, **QwenVL**, **InternVL** and **Gemini-1.5 Pro**.
- Developed and deployed a novel evaluation dataset comprising over **800** images and **two** distinct tasks to benchmark the performance of vision-language models. DATASET
- Collaborated with external research groups, notably **Google Research**, **IBM Research**, **Cohere**, and **Alibaba Research**, as well as interdisciplinary teams, to advance cutting-edge research in AI.

Data Science Fellow Jun 2023 – Jul 2023
Abu Dhabi National Oil Company (ADNOC - Panorama Department)

- Led and built an **ARIMA** time series model on **6.59 million** entries to forecast gas cracker flow rates, enhancing process efficiency. Performed end-to-end data preprocessing, cleaning, exploratory data analysis (EDA), feature engineering, and modeling.
- Engineered an NLP-based Q&A system for oil drilling reports using LLM APIs to extract insights and improve decision making. Experimented with embeddings from **DaVinci**, **GPT-3.5-turbo**, **Bard**, **Falcon-13B/40B** for query-context matching, evaluated on speed, cost, performance, and compute efficiency. Leveraged a domain-specific dataset of **74** daily oil drilling reports (PDFs) from **Utah FORGE**. CODE

Data Scientist Jul 2021 – Jul 2022
STAX Inc

- Conducted due diligence for private equity firms across **five** investment opportunities, leveraging data-driven insights and market analysis to evaluate potential deals and support strategic decision-making.
- Engineered pipelines to scrape **100K+** reviews and listings from major platforms, enabling in-depth analysis of market trends, competitors, and consumer sentiment to deliver actionable business intelligence for clients.
- Synthesized insights from **multiple** diverse data sources, including web scraping and survey data, to deliver strategic recommendations supporting clients' data-driven investment decisions.

Data Scientist Nov 2020 – Jun 2021
National Intensive Care Surveillance Unit (NICST)

- Led exploratory data analysis (EDA) and data transformation on clinical trial datasets comprising **96** variables for **800+** patients across **17** medical clinics, enabling analysis to support evidence-based decision-making in healthcare.
- Engineered automation scripts to streamline data mapping across **five** systems and formats, cutting the processing time of the previously implemented system by **half** and significantly reducing manual effort in clinical workflows.

Selected Projects

Label-free Adaptation of CLIP for Remote Sensing LINK Masters Thesis

- Engineered and deployed a label-free adaptation method for **Remote Sensing Scene Classification (RSSC)**, which outperformed the prior state-of-the-art by **5%** across **10** benchmarks.
- Explored the efficacy of auto-labelled prompt tuning by leveraging contextual knowledge from LLMs including **GPT-4o**, **LLAMA**, and **Gemini** to generate pseudo labels and adapt **CLIP** for remote sensing context.

Fine-Grained Image Classification Using Counterfactual Learning

- Proposed a **novel** method to explore the impact of learned and uncorrected (counterfactual) attentions on classification scores in fine-grained image classification.
- Conducted extensive experiments using **adversarially generated perturbations** on feature maps to generate counterfactual attentions, analyzing the effect of different perturbation combinations on model predictions.

Domain Adaptation for RGB to Thermal Images CODE

- Proposed a novel **Unsupervised Domain Adaptation (UDA)** approach for urban road scenes by transferring knowledge from RGB to thermal imagery using a triple-branch transformer architecture.
- Experimented with multiple transformer backbones (**DeiT**, **CvT**, **SWIN**) to classify pedestrians, cars, and bicycles across RGB and thermal domains; incorporated adversarial adaptation with a discriminator network and evaluated performance across various loss function combinations.

Football Game Outcome Prediction CODE

- Engineered and assessed the effectiveness of a machine learning models to predict football match outcomes using **35** player attributes, **13** team-level variables, and **11** previous match statistics.
- Scraped data for all **32 teams** participating in the FIFA 2021 World Cup using **BeautifulSoup** and **Selenium**, generating predictions from group stage matches to the grand finale.
- Experimented with various models including **tree-based classifiers**, **Gaussian Naive Bayes**, **regression models**, **SVM**, **XGBoost**, and **neural networks** to evaluate prediction performance.

Optimizing Direct Mail Fundraising

- Engineered a machine learning pipeline to optimize direct mail fundraising for a fictional organization using **8000+** data entries with **18** variables, including donation amount, neighborhood statistics, household demographics, and employment data.
- The pipeline included a classification model for predicting the likelihood of donation and a regression model for estimating the donation amount, experimenting with models such as **logistic regression**, **decision trees**, **random forests**, and **XGBoost**.

IoT Temperature Prediction with Dashboard

- Led the development of a **Node-RED** dashboard on **IBM Cloud** to visualize 12-month temperature forecasts based on **5000+** historical data points.
- Designed and implemented time series models, including **ARIMA** and **Prophet (Facebook)**, to predict temperatures in major Sri Lankan cities.

Visual Analytics with Batch and Streaming Data

- Deployed a mock hotel website with tracking codes to analyze user sessions via **Google Analytics** and **Data Studio**.
- Performed real-time streaming analytics on **190K+** Uber-Lyft data points using **Siddhi** and **MySQL**. Visualized insights using **Tableau** and **Power BI**, providing actionable insights for potential stakeholders.

Achievements

- Led a research team during the Undergraduate Research Internship Program (UGRIP) at MBZUAI, which won the prestigious **Best Team Award** for outstanding innovation and collaboration.
- Earned a competitive **fully-funded Master's scholarship** at MBZUAI, awarded for demonstrated academic excellence, leadership, and strong research potential in Artificial Intelligence.
- Consistently ranked among the top performers and featured on the **SLIIT Dean's List** for six consecutive semesters across the second, third, and final years.

Publications

For a complete list of my publications, please visit my Google Scholar profile.

References

- Available on request