# 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, NoSQL, NumPy, Pandas, SciPy, Scikit-learn, XGBoost, BeautifulSoup, Selenium, FastAPI, Flask, PyTorch, Keras
**AI Frameworks**: HuggingFace Transformers, MLflow, Weights & Biases, OpenAI API, LangChain, Agentic AI
**Data Engineering & MLOps Tools**: ETL, RAGs, Databricks, Docker, Airflow
**Cloud Platforms**: AWS (S3, EC2, SageMaker), Azure (ML Studio, Databricks), Google Cloud Platform
**Data Visualization Tools**: Tableau, Power BI, Seaborn, Matplotlib
**Development & Collaboration Tools**: Git, Bash scripting, Google Analytics

## Experience

**AI Researcher** — Jul 2023 – Present
*Mohamed Bin Zayed University of Artificial Intelligence*

- Led research on visual-temporal reasoning, developing the TemporalVQA benchmark to evaluate advanced MLLMs including **GPT-4o**, **LLAVA**, **LLAMA**, **InternVL**, **QwenVL** and **Gemini-1.5 Pro**, revealing significant limitations in temporal understanding and reasoning capabilities of MLLM.
- Contributed to **grassroots AI initiatives** aimed at enhancing cultural and linguistic diversity in AI systems, including the development of the CVQA Benchmark, a culturally-diverse multilingual VQA dataset covering **31** languages and **13** scripts across **30** countries, and the SEACrowd project, a multilingual multimodal data hub and benchmark suite designed to support **36** Southeast Asian languages.
- Collaborated with external research groups, notably **IBM Research**, **Cohere**, and **Alibaba Research**, as well as interdisciplinary teams, to advance cutting-edge research in AI.
- Served as a reviewer for ACL conferences

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

- Led and built an **SARIMAX** 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. 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 data transformation, EDA, and statistical modeling 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 — Master's 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.

**Optimizing Direct Mail Fundraising**

- Engineered a pipeline to optimize direct mail fundraising using **8,000+** data entries with **18** features, including donation history, demographics, and neighborhood statistics.
- Built a classification model to predict donor likelihood (**82% accuracy** and **0.79 AUC-ROC**) and a regression model to forecast donation amount (**RMSE of 14.2**).
- Experimented with **logistic regression**, **decision trees**, **random forests**, and **XGBoost**, selecting optimal models via **grid search** and **cross-validation**.

**IoT Temperature Prediction with Dashboard**

- Developed an interactive **Node-RED** dashboard on **IBM Cloud**, visualizing 12-month temperature forecasts across **50+ IoT sensor locations**.
- Implemented and compared time series models, including **ARIMA** and **Prophet (Facebook)**, achieving **MSE reduction of 18%** over baseline moving average models.

**Visual Analytics with Batch and Streaming Data**

- Deployed a hotel website with embedded tracking codes to monitor user behavior, leveraging **Google Analytics** and **Data Studio** to generate **daily traffic and conversion reports**.
- Simulated real-time analytics on **190K+** Uber-Lyft ride records using **Siddhi** and **MySQL**, identifying **peak-hour trends**, **geographic demand clusters**, and **fleet optimization opportunities**.
- Visualized batch and streaming insights in **Tableau** and **Power BI**, enabling data-driven recommendations for **marketing**, **operations**, and **resource allocation** strategies.

**Automating Automobile Accident Claim Process** LINK — Bachelors Thesis

- Spearheaded the design and implementation of an ML pipeline to automate the accident claim process, reducing manual handling time by **65%** through end-to-end computer vision and predictive analytics.
- Built a multi-stage pipeline integrating **vehicle make/model classification**, **damage localization via CNNs**, **repair cost estimation using regression models**, and **customer churn prediction** based on tabular and image-based data.

**Football Game Outcome Prediction** CODE

- Developed machine learning models to predict FIFA World Cup 2022 match outcomes using **35 player-level attributes**, **13 team statistics**, and **11 historical match features**.
- Scraped and compiled structured data for all **32 national teams** using **BeautifulSoup** and **Selenium**, automating coverage from group stages to the finals.
- Evaluated multiple algorithms including **tree-based classifiers**, **Naive Bayes**, **SVM**, **XGBoost**, and **neural networks**, achieving up to **67% accuracy** on validation sets.

## 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.

## Publications

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

## References

Available on request