

Mr. Fakruddin Mohammed
Mohammed_fakruddin@yahoo.co.uk
+44 7809 436 295 (UK Citizen)

AI Projects Profile

Language Identification

(2019)

Deep Learning for Foreign Language Identification: Developed a convolutional neural network (CNN) model to accurately identify foreign languages in text data. The project involved training the model on a diverse dataset of multilingual text samples, enabling it to learn distinct language features and patterns. Implemented using Python and TensorFlow, the model achieved high accuracy in classifying languages, demonstrating its potential for applications in multilingual content processing and language detection systems.

(<https://github.com/mohammed-fakruddin/DeepLearningCNNLanguageIdentification>)

Autonomous Car – Vision/CNN Neural Networks

(2019)

Deep Learning for Autonomous Car Navigation: Developed a deep learning model to enable autonomous navigation in robotic vehicles. The project involved training the model using simulated driving environments to perform tasks such as lane following, obstacle avoidance, and path planning. Techniques like convolutional neural networks (CNNs) were employed to process visual inputs and make real-time driving decisions. The trained model was tested in various simulated scenarios to evaluate its performance and robustness.

(<https://github.com/mohammed-fakruddin/DeepLearningAutonomousCarNavigation>)

Boson Particles Prediction

(2020)

Machine Learning for Higgs Boson Detection in Particle Physics: Developed a machine learning model to enhance the detection of the Higgs boson in high-energy physics experiments. Utilizing advanced algorithms, the model analyzes particle collision data to identify potential Higgs boson events, improving detection accuracy and contributing to a deeper understanding of fundamental particles.

(<https://github.com/mohammed-fakruddin/MachineLearningForParticlePhysicsHigBoson>)

Fund Search – Elastic Search, UX, UI

(2020)

Rank-Based Search Engine for Investment Funds: Developed a search engine tailored for the investment fund industry, utilizing ranking algorithms to enhance the retrieval of relevant fund information. The system indexes fund data and applies ranking

methodologies to present users with the most pertinent results based on their queries. This approach improves information accessibility for investors and financial analysts, facilitating more informed decision-making.

(<https://github.com/mohammed-fakruddin/RankBasedSearchEngineForFundIndustry>)

Semantic Search – Fund Objectives (2020)

NLP-Based Linguistic Search Engine for Fund Objectives: Developed an AI-powered search engine that bridges the gap between legal financial language and retail investor queries using advanced NLP algorithms. Implemented using Python Flask, NLTK, Spacy, Gensim, TensorFlow, and Scikit-Learn, with data extracted via BeautifulSoup from 3000+ mutual fund prospectuses. Utilized NLP models like Word2Vec, TF-IDF, ELMO, Universal Sentence Encoder, and Fast Text to evaluate semantic similarity and recommend top-matching funds. Deployed on AWS EC2 for real-time querying and result generation.

(<https://github.com/mohammed-fakruddin/NLPLinguisticSearchEngineForFundIndustry>)

PDF Comparison (2021)

Fund Documents Comparison Using Robotics and Computer Vision: Designed a robust PDF comparison solution for the fund industry using computer vision and robotics techniques to detect and highlight document changes. Implemented with Python, OpenCV, and Matplotlib, the system extracts 40+ metadata attributes (e.g., GPS, camera, and scanner details) and assigns probabilistic confidence scores to changes, overcoming limitations of timestamp-based methods.

(<https://github.com/mohammed-fakruddin/FundDocumentsComparisonUsingRoboticsAndComputerVision>)

Digital Assistant for Fund Specialists – Voice bot (2021)

Fund Navigator: NLP Digital Assistant for Fund Specialists

Developed an AI-powered digital assistant, FundNavigator, to enhance the efficiency of fund specialists in managing investment portfolios. Utilizing natural language processing (NLP) techniques, the assistant interprets user queries related to fund performance, risk assessments, and market trends, providing accurate and timely insights. Implemented with Python and integrated with financial data APIs, FundNavigator streamlines information retrieval, enabling fund specialists to make informed decisions swiftly.

(<https://github.com/mohammed-fakruddin/NLPDigitalAssistantForFundSpecialists>)

AI for Astronomy

(2021)

NLP Digital Assistant for Astronomy and Telescope Control: Developed an AI-powered digital assistant designed to enhance the efficiency of astronomers by integrating natural language processing (NLP) capabilities with telescope control systems. This assistant enables users to operate telescopes through voice commands, facilitating tasks such as object tracking, data acquisition, and real-time adjustments. Implemented using Python and integrated with astronomical databases, the system streamlines observational workflows, allowing astronomers to focus more on data analysis and discovery.

Deep Learning - AutoDrone Valet Parking System

(2021)

Developed an autonomous drone-based valet parking system designed to alleviate parking challenges in congested urban areas. Utilizing advanced computer vision and machine learning algorithms, the drone navigates parking facilities, identifies available spaces, and guides vehicles to optimal parking spots. The system enhances parking efficiency and reduces the time drivers spend searching for parking, contributing to improved urban mobility.

<https://github.com/mohammed-fakruddin/AutoDroneValetParking/tree/main>

Data Discovery – Chatbot (RASA Framework)

(2022)

Designed and implemented a self-service data discovery tool at LGIM to address challenges in navigating a 49TB data lake with 125+ databases, 3500+ tables, and 7.5M attributes. Built using Python, Flask, Cloudera Data Platform, Cloudera Machine Learning, and Domino Cloud, the solution leveraged a RASA-based chatbot for BM25 semantic search, text-to-SQL data observability, ESG data queries, and general Q&A, enabling business users to efficiently locate and utilize data within the data lake platform.

(https://zarapiano.s3.us-east-1.amazonaws.com/LGIM_CHAT_MPT_V7.mp4)

Data Discovery – Metadata – Knowledge Graphs

(2022)

Knowledge Graphs for Data Discovery and Lineage: Developed a knowledge graph framework to enhance data discovery, lineage, and governance for financial datasets. Implemented using semantic metadata, ontologies, and RDF triple stores to answer critical business questions on data usage, governance, and impact analysis. Demonstrated applications in regulatory reporting, operational risk management, and cost analysis across financial institutions, drawing from the concept of the Semantic Web.

Data Tagging – NLP Topic Modelling – LDA/LSA (2023)

Developed and implemented an NLP-based topic modelling solution at LGIM to enhance data discovery within a 49TB data lake containing 125+ databases, 3500+ tables, and 750K data fields. Utilizing LDA and LSA algorithms, the project categorized all data attributes into 30 distinct topics, enabling users to refine their searches by selecting relevant topics, significantly improving the accuracy of semantic search. This solution greatly enhanced user experience and received widespread appreciation from business users for its effectiveness in simplifying data access.

Structured Data from Unstructured Text (2024)

Developed an end-to-end pipeline at LGIM to extract quantitative data from nearly 500,000 unstructured PDF Investment Mandate Agreement (IMA) documents containing critical information such as inflation bases, projected cashflows, discount rates, trigger levels, investor details, and permitted/prohibited investment instruments. Implemented using Python, Flask, Azure OpenAI, and OCR technologies, the solution successfully transformed unstructured PDF text into structured data, enhancing the management of pension investments and streamlining data accessibility.

AI Agents – Text 2 SQL (2024)

Developed an AI-powered Text-to-SQL chatbot for LGIM to query trade, risk, PnL, ratings, market data, and reference data from internal and external sources on the Cloudera Data Lake platform. Implemented using PandasAI, Langchain, Langgraph, and CrewAI agent frameworks with OpenAI models, the project initially faced performance challenges with 3rd Normal Form databases but achieved significant improvements when data was modelled using a star schema. This experimental project is ongoing, aiming to streamline data retrieval through natural language queries.

PowerPoint Generation (2024)

Automated the generation of quarterly LGIM fund performance reports by developing a Python-based pipeline using OpenAI, Pandas AI, Crew AI, and Langchain agents. The solution integrated three AI agents to extract fund performance data, and benchmark indices, and aggregate the information into PDF and PowerPoint documents adhering to LGIM's strict marketing and branding standards for colours, fonts, and formats. This automation significantly enhanced productivity for the marketing and client relationship teams and was well received by the Index Trading team as a successful proof of concept.

Genetic Algorithms for Prediction (2024)

Developed a research study on employment forecasting in India's labour market, testing the applicability of Phillips curve and Okun's law, which traditionally predict

unemployment through inverse relationships with inflation and GDP. Recognizing the unique challenges in low/middle-income economies like India, where the labour force is predominantly employed in agriculture and seasonal jobs, the study employed genetic algorithms (GA) for feature selection. Results indicated that the Phillips curve did not hold well for India's labour market during the studied period, while GA-based feature selection identified economic growth, household consumption, government spending, and coal imports as key macroeconomic variables influencing unemployment.

<https://doi.org/10.59720/22-252>

Companion Robot

(2025)

An AI and LoRa-enabled companion robot offers an effective remote health and safety surveillance solution for ageing individuals living alone in rural areas, addressing the challenges posed by population ageing and rural-to-urban migration. Unlike systems plagued by false alarms, unreliable internet, and lack of emotional care, this system integrates AI for real-time health monitoring, emergency support, and emotional companionship. It consists of a C1001 mmWave sensor for continuous health and activity data collection, a Raspberry Pi-based AI edge device with a locally deployed large language model (LLM) for processing data, voice-based anomaly validation, and emergency response triggering, and a LoRa-based remote assistance dispatch unit for low-power, long-range emergency alerts. Key advantages include AI-driven anomaly validation, independence from internet connectivity, and AI-based social interaction. Experimental tests show reliable health anomaly detection, real-time operation, and prompt emergency assistance, providing a privacy-conscious, AI-powered elderly care solution, with future research focusing on enhancing AI decision-making, energy efficiency, and predictive healthcare analytics.

Indian Sign Language

(2025)

Developed an inclusive bidirectional Indian Sign Language (ISL) translation system enabling real-time sign-to-text and text-to-sign communication. Utilizing Generative AI, Deep Learning, CNNs, LSTMs, and Google MediaPipe, the system ensures seamless interaction between ISL users and non-signers. A lookup-based method with pre-defined sign poses minimizes computational needs and expands vocabulary from 263 to over 2900 words. GPT-4o aids sign-to-text translation with 53% perfect similarity and 75% scoring above 60%, enhancing accessibility for India's diverse, underrepresented communities.