

Mohammad Faseeh Ahmed

mm9314@g.rit.edu | +1 585 202 5217 | [LinkedIn](#) | [Github](#) | [Portfolio](#) | [Kaggle](#) | [Tableau](#)

EDUCATION

Rochester Institute of Technology, Rochester, NY, M.S in Data Science

Expected May 2025

Coursework: Neural Networks, Software Engineering for Data Science, Applied Statistics.

GPA: 3.84/4.00

Jawaharlal Nehru Technological University Hyderabad, B.Tech in Computer Science

July 2018 - July 2022

Coursework: Data Structures and Algorithms, Computer Vision, Artificial Intelligence, NLP

GPA: 3.2/4.00

SKILLS

Programming Languages: Java, Python, C++, R, JavaScript, Object Oriented Programming(Python, Java)

Frameworks: PyTorch, Keras, Scikit, Tensorflow, Groovy, PySpark, Flask, React, NodeJS,

Databases: SQL, MongoDB, SQLite, MySQL, NoSQL, PostgreSQL, DynamoDB

Technologies/Tools: JSON, Docker, Git, AWS, Kafka, GitLab, Numpy, Pandas, MLflow, Postman, Tableau, Power BI, MLOps

ML Algorithms and Techniques: Regression, Classification, Clustering, Recommender Systems, Deep Learning, NLP, A/B

Testing, Time Series, Optimization, EDA, ETL,

PROFESSIONAL EXPERIENCE AND INTERNSHIPS

Daiichi Sankyo Inc, Basking Ridge, NJ- R&D Data Governance Intern

03/2024 - Present

- Developed an Informed Consent Forms (ICF) analysis tool using BERT and T5 models, also testing Amazon Bedrock for large-scale language processing.
- Built a Flask-based frontend for the ICF tool, enabling secure modifications of document classifications based on user permissions.
- Implemented advanced Large Language Models (LLMs) using Amazon Bedrock, achieving a 20% increase in classification accuracy for detecting explicit prohibitions against data sharing.
- Utilized Amazon SageMaker for model training and fine-tuning, streamlining the handling of vast volumes of legal documents to ensure scalability and efficiency.
- Optimized model inference and processing times with EC2 instances, improving overall performance and reliability of the ICF analysis tool.
- 1. Designed an on-premise chatbot delivering instant trial data, cutting access time by 70%.
- 2. Enabled visualizations in the chatbot, improving data comprehension and decision-making efficiency.
- 3. Automated milestone and site tracking, enhancing study performance and reducing missed deadlines by 30%.
- 4. Developed a secure chatbot architecture, ensuring 100% compliance with company data regulations.
- 5. Integrated RAG and AI, achieving 80% improvement in answer accuracy and user trust.
- Led the migration of clinical data from Veeva Vault to Redshift, establishing an ETL pipeline that ensured data accuracy and improved analytics accessibility.
- Built indexing and foreign key relationships in Redshift to optimize query performance, enabling high-speed analysis on clinical data.
- Implemented data validation checks during migration to uphold data governance and regulatory compliance, ensuring reliable and accurate data in the new environment.
- Collaborating with stakeholders to align the tool's capabilities with Daiichi Sankyo's data-sharing policies, compliance standards, and real-world data requirements.

Rochester Institute Of Technology, Rochester, NY- Research Assistant

08/2024 - Present

- Collaborating with Professor Haibo Yang to enhance federated learning models, utilizing gRPC and PyTorch to implement scalable decentralized training algorithms.
- Analyzing and optimizing distributed environments with PyTorch RPC, reducing communication overhead by 15% through improved data transfer protocols.
- Leading the integration of the FedDisco algorithm into existing frameworks, accelerating model convergence and reducing training time by 20%.
- Transitioning between distributed communication algorithms, leveraging FedDisco's adaptive communication strategies for efficient, real-time model updates.
- Developing advanced distributed training techniques for Large Language Models (LLMs), achieving a 30% increase in system performance and minimizing network latency.
- Utilizing gRPC for dynamic, secure communication within federated learning models, enhancing data synchronization and increasing overall model accuracy by 10%.
- Designing hands-on projects for the Intro to Machine Learning course, applying theory to real-world machine learning scenarios using Python and PyTorch.
- Providing personalized support to over 25 students, fostering problem-solving skills and guiding successful project outcomes, significantly improving their understanding of distributed machine learning concepts.

SEO Content AI, Los Angeles, CA - AI Infrastructure Engineer

Nov 2022 - July 2023

- Enhanced AI-driven content generation efficiency by 25% through leading the integration of transformers within AWS microservices.
- Developed a Chrome extension using Python, JavaScript, and NodeJS, boosting content quality and generation speed by 40%.
- Utilized Docker for scalable and reliable deployment across cloud environments via AWS ECS and Fargate.
- Improved project management and team collaboration through implementing effective development methodologies.

- Optimized content creation for various use cases by conducting A/B testing with models like LLaMA-13B, Flan-T5, Vicuna-13B, and GPT-3.5.
- Broadened audience reach by using GCP's Google Translate API for multi-language content translation.
- Ensured high-quality, error-free content by implementing NLP algorithms with BERT, RoBERTa, and DistilBERT for grammar correction, significantly enhancing user trust.

White Label Resell, Los Angeles, CA - Machine Learning Engineer

June 2022 - March 2023

- Automated article generation for digital marketing using AWS Lambda, NodeJS, and API Gateway achieving a 60x reduction in operational costs.
- Integrated NLP and TensorFlow to analyze and generate over 130K high-quality articles within a week, improving content strategy.
- Utilized DynamoDB for managing large datasets, ensuring efficient data access and manipulation for content generation processes.
- Fine-tuned advanced NLP models like BERT, RoBERTa, ALBERT, and DistilBERT for contextual understanding, improving the relevance and quality of generated articles.
- Explored generative models including GPT-3, GPT-Neo, GPT-J, and GPT-NeoX, to diversify content creation, ensuring a wide range of engaging and unique articles.
- Leveraged Flan-T5 and T5 Codegen for specific content generation tasks, optimizing for both efficiency and creativity in article production.
- Developed custom dataset models with Gensim, enhancing thematic accuracy and alignment with marketing goals.
- Employed MLOps practices with Git and Docker to streamline the development, training, and deployment of machine learning models, facilitating continuous improvement and collaboration.

Digital Clinics Research and Services, Hyderabad, India - Data Scientist Intern

Nov 2021 - Dec 2022

- Developed a sophisticated image classification system using Faster R-CNN, tailored for the nuanced detection of cancerous cells in histopathological images, leveraging TensorFlow for model training and optimization.
- Engineered a custom segmentation solution with Detectron2, enabling precise delineation of tumor boundaries in medical scans. This approach utilized QuPath for image management and Groovy for scripting complex analysis workflows, significantly enhancing the quality of medical image analysis.
- Implemented YOLOv5 for the automated screening of pathological slides, achieving unprecedented speed and accuracy in identifying critical diagnostic markers, integrated with OpenCV for real-time image processing and enhancements.
- Led the integration of MLOps practices, utilizing Docker for containerization and AWS for model deployment, ensuring scalable and robust AI solutions in clinical settings.
- Mentored a diverse group of 10-20 junior colleagues (first and second-year students) on multiple ML and DL projects, fostering a collaborative learning environment and enhancing the team's overall AI expertise and project execution capabilities.

Edgeforce Solutions, Hyderabad, India - Data Scientist Intern

Nov 2021 - Feb 2022

- Built a YOLOv5-based real-time object detection system using Python and TensorFlow, integrated with Streamlit for a user-friendly interface, achieving 90% accuracy.
- Engineered a deep learning model for speech recognition with PyTorch, achieving 95% accuracy, deployed in an Army Walkie Talkie emulator.
- Utilized Docker and AWS for deploying scalable and efficient machine learning models, reducing operational costs and improving resource utilization.
- Leveraged SQL and SQLite for efficient data storage and retrieval, enhancing system performance and data management.
- Conducted A/B testing to optimize the model's performance, resulting in enhanced accuracy and user satisfaction.

PROJECTS

Chronic Kidney Disease Predictor

Nov 2022 - July 20

- Utilized Python's Pandas and Numpy for data manipulation and preprocessing, ensuring accurate model inputs.
- Implemented Scikit-learn to develop a logistic regression model, achieving a 98% F1 score for kidney disease prediction.
- Orchestrated a hybrid application structure, employing Flask for backend functionalities and leveraging the MERN stack for frontend development, to provide an intuitive platform for disease prediction.
- Employed Matplotlib and Seaborn for data visualization, providing insightful analytics for model performance evaluation.
- Applied MongoDB for efficient data storage, enabling scalable and secure management of patient data.

Beverage Management System

June 2022 - March 2023

- Implemented TensorFlow for creating a convolutional neural network model, achieving real-time beverage detection with 98% accuracy.
- Used OpenCV for image processing, enhancing the model's ability to recognize various beverages under different conditions.
- Developed a web application using Flask as the backend and ReactJS for the frontend, offering a seamless user experience.
- Leveraged PostgreSQL for database management, ensuring robust and reliable storage of inventory data.

- Incorporated Docker for deploying the application, ensuring consistency across development and production environments.

Covid19 Bot

Nov 2022 - July 20

- Developed a chatbot using Python, Flask, and DialogFlow to offer real-time COVID-19 updates via Telegram, enhancing accessibility for users.
- Implemented Pandas and Numpy for robust data analysis and numerical computations to provide accurate and up-to-date health statistics.
- Integrated MongoDB to efficiently store and manage user queries and responses, facilitating personalized interaction and data retrieval.
- Utilized RapidAPI for accessing live COVID-19 data, ensuring the chatbot's information was current and reliable for users seeking instant updates.

Flight Price Predictor

Nov 2022 - July 20

- Enhanced the predictive model using Scikit-learn's regression techniques and LSTM networks for dynamic price forecasting, incorporating temporal data analysis.
- Conducted extensive feature engineering with Pandas, improving model accuracy by identifying key price determinants from historical flight data.
- Utilized Matplotlib and Seaborn for insightful visualizations of data trends and model performance, aiding in the interpretability of predictions.
- Deployed the enhanced model on Heroku with a Flask web application interface, providing users with real-time, accurate flight price predictions.

Generating Synthetic Anime Data

June 2022 - March 2023

- Leveraged PyTorch and a DC GAN architecture to create high-quality synthetic anime images, significantly enhancing the diversity of our machine learning datasets.
- Utilized OpenCV for advanced image preprocessing and augmentation tasks, including resizing, normalization, and color adjustments, to prepare images for training, enhancing the model's ability to generate diverse and realistic anime characters.
- Developed Python scripts to automate the collection and preprocessing of anime images, streamlining the dataset creation process and enabling the efficient generation of large volumes of training data.
- Applied OpenCV's edge detection and image filtering capabilities to enhance the detail and quality of generated images, ensuring a high degree of realism and variability in the synthetic data.
- Integrated additional data augmentation techniques such as flipping, rotation, and scaling through OpenCV, further augmenting the training dataset to improve the robustness and generalization of the GAN model.

Number Plate Detection

Nov 2022 - July 20

- Deployed YOLOv5 for real-time number plate detection, achieving high accuracy in diverse lighting and environmental conditions.
- Incorporated OpenCV for advanced image processing tasks, such as adaptive thresholding and contour detection, to improve the accuracy of number plate recognition.
- Developed a system using Flask as the backend, enabling real-time detection and information processing, with a user-friendly interface for monitoring and alerts.
- Integrated OpenCV's Gaussian Blur to reduce image noise and improve the model's detection capabilities in varying environmental conditions.
- Implemented perspective transformation techniques in OpenCV to correct angles and enhance the readability of number plates, ensuring accurate recognition regardless of the vehicle's orientation.

Research Papers

Tubules Detection on Breast Carcinoma Whole Slide Images Using Artificial Intelligence (Sep 2022):

- This groundbreaking research utilized advanced AI deep learning models to accurately detect tubules in breast carcinoma images. The abstract was recognized for its innovation and selected for presentation at the prestigious International C-MIMI Conference at Johns Hopkins University, underscoring the project's contribution to medical imaging and diagnostics.

Web-Based Mitosis Detection on Breast Cancer Whole Slide Images Using FasterRCNN and YOLOv5 (Dec 2022):

- This significant paper, published by the Science and Information (SAI) Organization in the International Journal of Advanced Computer Science Applications (IJACA), detailed the development and application of Faster R-CNN and YOLOv5 for the detection of mitosis in breast cancer whole slide images. The research provided insights into the capabilities of these models in enhancing the accuracy and efficiency of breast cancer grading, demonstrating a pivotal step forward in the application of machine learning in healthcare.

ACHIEVEMENTS

- **Kaggle Notebook and Dataset Expert**
- **Open Source Contributor at OpenVino(Intel), Ivy**
- **Won Best Overall Hack in CSH Hacks Hackathon 2023** : Developed a Chrome extension with a streamlined

front-end, leveraging AWS Lambda for backend deployment, which integrates OpenAI's GPT-3.5 to scrutinize terms and conditions, enhancing user vigilance through AWS-powered processing and API Gateway for efficient data handling, encapsulating cloud architecture expertise in delivering a user-centric solution.