Mohammed Mahyoub

+1 (607) 727-3208 — mahyoubds@gmail.com — linkedin.com/in/mohammedamahyoub — github.com/mmahyoub

SUMMARY

Highly accomplished and results-driven Data Scientist with a Ph.D. in Industrial and Systems Engineering, specializing in the application of Large Language Models (LLMs) and machine learning to revolutionize healthcare analytics. Proven expertise in leading advanced analytics initiatives, developing and deploying impactful predictive models, and significantly enhancing clinical decision-making and operational efficiencies. Adept at translating complex data into actionable insights and fostering innovation within EMR systems and cloud environments.

EDUCATION

Binghamton University — Binghamton, New York
Ph.D. in Industrial and Systems Engineering Aug 2020 – Apr 2025
Dissertation: *From Clinical Text to Informed Decisions: A Study of Large Language Models in Radiology*

Binghamton University — Binghamton, New York
MS in Industrial and Systems Engineering Aug 2018 – Aug 2020

Jordan University of Science and Technology — Irbid, Jordan
BS in Biomedical Engineering Sep 2013 – Aug 2018

TECHNICAL SKILLS

Core Technologies: Predictive Modeling, Machine Learning, Deep Learning, Natural Language Processing (NLP), Large Language Models (LLMs), Computer Vision, MLOps, Statistical Modeling, Simulation Modeling, Mathematical Modeling, Computational Thinking, Software Development, Transfer Learning, LLM Fine-tuning, Retrieval Augmented Generation (RAG), Agentic Workflows, Advanced Analytics.

EMR Systems: Epic, Epic Clarity, Epic Predictive Model Administration, Epic Slate (Nebula).

Programming Languages: Python, SQL, Mathematica

Tools, Libraries, and Frameworks: Scikit-Learn, PyTorch, Langchain, Transformers, Hugging Face, CrewAI, Pandas, Numpy, Matplotlib, XGBoost, Git, GitHub, Docker, CI/CD, PowerBI, VS Code.

Cloud Computing: Azure (Azure Container Registry, Azure App Services, Azure Functions, Azure Machine Learning Studio, Azure AI Studio, Azure CLI, Azure Python SDK), Fabric, Azure Synapse Analytics.

WORK EXPERIENCE

Data Scientist Manager — Virtua Health, Marlton, NJ Jun 2023 – Present
(80% Hands-on; 20% Managerial)

- Spearheaded and managed 3+ advanced analytics initiatives, delivering Power BI dashboards that significantly enhanced the monitoring of ICU, neurology, and heart failure quality metrics, leading to more informed clinical interventions and improved resource allocation.
- Directed the end-to-end development and deployment of 3+ critical predictive models, including:
  - Level of Care Model: Achieved a 20% reduction in claims denials.
  - Hospital-at-Home Eligibility Model: Realized $> 70\%$ accuracy in identifying suitable candidates.
  - Heart Failure DRG Classification Model: Attained 86% accuracy, optimizing resource allocation.
- Engineered and deployed a pioneering LLM-based stroke abstraction system, slashing patient chart review time by over 70% (from 1.5 hours to < 25 minutes).
- Developed and launched an innovative LLM-powered HR chatbot web application, dramatically improving information retrieval speed and reducing manual search time from about 5 minutes to < 1 minute.

- Engineered and deployed an advanced hybrid-search enabled chatbot for the access center, designed to significantly improve information retrieval speed and accuracy; currently in the final validation phase prior to wider deployment.
- Authored three peer-reviewed publications in high-impact journals, showcasing successful applications of machine learning and LLMs in healthcare.
- Architected, developed, and deployed a comprehensive ML fairness evaluation tool to systematically assess and mitigate algorithmic bias in predictive models, promoting equitable healthcare delivery.
- Pioneered MLOps training programs, conducting 5+ workshops to empower the team in deploying models within EMR systems and on Azure Cloud.
- Leveraged LLMs to architect high-performance key insight extractors from clinical notes for 4+ initiatives, notably extracting CRT documentation for cardiology patients, improving data accuracy and accessibility.

Data Scientist — Virtua Health, Marlton, NJ Mar 2022 – Jun 2023

- Developed and deployed a patient discharge date predictive model achieving 75% accuracy, outperforming the commercial baseline and contributing to over 26,000 excess bed-days saved since 2023.
- Engineered and implemented a sepsis predictive model with 93% sensitivity, drastically reducing the false positive rate from 30% (commercial model) to 6%, enhancing patient safety.
- Created and deployed an LLM-based pulmonary embolism (PE) text classification algorithm for radiology reports, achieving $> 99\%$ AUC, significantly improving data availability and identifying over 700 positive cases in 2024.
- Established robust cloud computing pipelines for scalable training and deployment of machine learning models using Azure and Epic Nebula.
- Developed and operationalized advanced NLP modules to extract structured data and actionable insights from physician notes with exceptional accuracy ($> 90\%$), directly contributing to improved data quality for clinical research.

Graduate Research Associate (Data Scientist) — Watson Institute for Systems Excellence (WISE), Binghamton University, Binghamton, NY Jan 2019 – Mar 2022

Research Project Location: Montefiore Health System CMO, Yonkers, NY

- Developed, deployed, and maintained staff and investment monthly time reporting systems for six departments, significantly enhancing time management and resource optimization.
- Designed and managed dynamic data dashboards for 3+ care management programs, providing critical insights for program evaluation and improvement.
- Conducted advanced analytics to identify and address critical barriers in the patient discharge process, recommending data-driven solutions.
- Led 2+ impactful healthcare process improvement projects, including the optimization of the referral processing unit, resulting in streamlined workflows.
- Developed an innovative machine learning-guided simulation framework to streamline health referral processing, forming the basis of MS thesis.
- Researched and synthesized literature to prepare comprehensive training materials for the Process Innovation and Engineering Department.
- Authored research publications demonstrating novel applications of machine learning, simulation modeling, and process improvement frameworks to solve complex healthcare challenges.

CERTIFICATIONS

- Azure Data Science Associate, Microsoft (05/2023)
- Deep Learning Specialization, DeepLearning.AI (04/2023)
- Machine Learning Specialization, DeepLearning.AI (10/2022)
- Clarity Data Model (ID: 335035212), Epic (10/2022)

- Microsoft Azure Fundamentals (ID: 993833094), Microsoft (09/2022)
- Caboodle Data Model (ID: 335035211), Epic (09/2022)
- Cogito (ID: 335035213), Epic (09/2022)
- Design Thinking for Healthcare, Binghamton University (07/2020)

SELECTED PUBLICATIONS

Access full list on Google Scholar

- Mahyoub, M., Wang, Y., & Khasawneh, M. T. (2025). Automating radiology impressions: Enhancing accuracy and efficiency with fine-tuned Mistral 7B LLM. *Neural Computing and Applications.* (Under review)
- Mahyoub, M., Wang, Y., & Khasawneh, M. T. (2025). Identification and extraction of follow-up recommendations from radiology reports using a fine-tuned LLM. *Natural Language Processing.* (Under review)
- Mahyoub, M., Wang, Y., & Khasawneh, M. T. (2025). GPT-4o in radiology: In-context learning based automatic generation of radiology impressions. *Natural Language Processing Journal, 100145.*
- Mahyoub, M., Dougherty, K., & Shukla, A. (2025). Extracting Pulmonary Embolism Diagnoses From Radiology Impressions Using GPT-4o: Large Language Model Evaluation Study. *JMIR Medical Informatics, 13*(1), e67706.
- Mahyoub, M. A. (2025). Integrating multi-criteria decision making and simulation modelling to improve health referral processing. *International Journal of Management and Decision Making, 24*(2), 107-131.
- Mahyoub, M. A., Dougherty, K., Yadav, R. R., Berio-Dorta, R., & Shukla, A. (2024). Development and validation of a machine learning model integrated with the clinical workflow for inpatient discharge date prediction. *Frontiers in Digital Health, 6*, 1455446.
- Mahyoub, M. A., Yadav, R. R., Dougherty, K., & Shukla, A. (2023). Development and validation of a machine learning model integrated with the clinical workflow for early detection of sepsis. *Frontiers in Medicine, 10*, 1284081.

AWARDS AND HONORS

- IISE Student Innovative Design Competition, First Place (2020)
- Alpha Pi Mu, Industrial Engineering Honor Society (2020)
- Research Assistantship and Tuition Scholarship from Binghamton University (2018)
- Governmental Full Scholarship for Bachelor's Studies (2013-2018)