MARK FAHIM

(561) 662-1835 • mark.fahim50@gmail.com • linkedin.com/in/mark-G-fahim • mark-fahim.vercel.app

---

CORE SKILLS

Machine Learning, NLP & LLMs: Agentic AI, Conversational RAG, Large Language Models (LLMs), Fine-tuning & Alignment (DPO, GRPO), Embeddings & Vector Search, Deep Learning, Natural Language Processing (NLP), Model Evaluation, Generative AI, XGBoost, Keras, TensorFlow, Scikit-learn.

Data Engineering & MLOps: Data Wrangling & Preprocessing (Chunking/Loaders), SQL, ETL Pipelines, Data Ethics, Dockerized Model Services, HPC (ROAR), Experiment Tracking, PostgreSQL, SQL Server.

Back-end & Cloud Development: NestJS, FastAPI, ASP.NET Core, RESTful APIs, AWS (Lightsail), Azure, Firebase, Authentication & Logging, Scalable Chat Backends.

Analytics & Visualization: Statistical Analysis, Experiment Design (A/B Testing), Tableau, Power BI, Python (Pandas, NumPy, Matplotlib, Seaborn), Insight Communication.

Front-end & Mobile: Next.js, Android (Java, Kotlin), Flutter (Dart), Git/GitHub, Performance Optimization.

Methods & Collaboration: Agile (Scrum Master, Sprint Planning), Cross-Functional Collaboration, Technical Documentation, Requirements Refinement.

WORK EXPERIENCE

Pennsylvania State University, University Park, PA May, 2024 – Present

AI Research Assistant

- Built machine learning models using TensorFlow and Scikit-learn to predict opioid-related deaths. Improved model accuracy by 15% through feature selection and hyperparameter tuning, helping inform public health policy decisions.
- Led the Sustainable Energy Marketplace project, analyzing public sentiment using NLP techniques. Identified key correlations between societal support, investment trends, and job creation. Presented findings at Massey'24 and ICDS Symposium'24; additional submissions under review.
- Designed web-scraping pipelines using Python, BeautifulSoup, and Scrapy to collect documents for an NSF-funded research study on gender disparities in bankruptcy. Processed over 5 million records and uncovered socioeconomic patterns across different demographics.
- Developed and fine-tuned large language models with Retrieval-Augmented Generation (RAG) using reinforcement learning techniques. Created an AI evaluation framework with human feedback to improve model quality and reduce errors in chatbot responses.
- Deployed predictive models to production on Penn State's HPC ROAR infrastructure using Docker and AWS. Ensured models were scalable, reproducible, and accessible to multiple research teams across campus.

ReviewBiz, West Palm Beach,FL, USA March, 2023 – April, 2024

Lead Software Engineer & Scrum Master

- Led team through Agile sprints, improving team velocity and on-time delivery by 20%. Managed sprint planning, daily standups, and retrospectives to keep projects on track and address roadblocks.
- Built REST APIs using NestJS and FastAPI to handle thousands of daily requests. Added security features, logging, and error handling to keep systems running smoothly with minimal downtime.
- Created automated data validation dashboards using Tableau and Python. Reduced manual data checking work by 80%, saving hours each week and helping business teams catch errors faster.

- Led the migration from NoSQL to SQL databases, redesigning the data structure to support more complex queries. Improved data retrieval speed by 25% and trained junior engineers on database best practices.
- Worked with product, design, and operations teams to align on priorities and remove blockers. Ran effective retrospectives that led to process improvements and stronger team collaboration.

Integrated Diagnostics Holding (IDH), Giza, Egypt Feb 2021 – Sept 2022

Software Engineer – Data & Machine Learning Specialist

- Managed data migration for 36 million annual diagnostic records. Built SQL validation scripts to find and fix duplicate records, incomplete information, and data quality issues. Ensured clean data for all 600+ branches nationwide.
- Built automated workflows in Power Automate and SharePoint to manage data access and approvals. Different branches could see only their own data while headquarters maintained oversight, improving compliance and data security.
- Designed a document management system with automated approval workflows, alerts, and notifications. Reduced time to retrieve documents by 30% and made it easier for teams to follow compliance rules.
- Created Power BI dashboards to track data quality across all branches. Identified and fixed patterns that were causing errors, reducing reporting mistakes by 25%.
- Worked with IT, compliance, and operations teams to write data governance policies that met HIPAA requirements. Documented best practices to help teams handle patient data securely during the COVID-19 response.

EDUCATION

Pennsylvania State University, University Park, PA, USA Dec 2025
Master's Degree in Artificial Intelligence

Penn State Great Valley School, Malvren, PA, USA Sept 2025
Graduate Certificate in AI Engineering**,** Natural Language Processing (NLP)

Cairo University, Faculty of Computers and Artificial Intelligence, Cairo, Egypt Aug 2020
Bachelor's Degree in Computer Science and Information Systems

---

PUBLICATIONS & RESEARCH CONTRIBUTIONS

- Sustainable Energy Marketplace and Societal Readiness, with Hyungjin Kim, Saya Lee, and Darshana Sunoj — Presented at Massey'24 and ICDS Symposium'24 (Under review: SWFA'25, ANS'25, WEAI'25, GFC'25, MRS'25)
- Social Readiness Level and Natural Language Processing: Tools for Measurement and Management of Societal Acceptance of Nuclear Energy, with Hyungjin Kim, Saya Lee, and Darshana Sunoj (Accepted: ANS'25)
- The Impact of Tax Law Changes on Bank Payout and Performance, with Rebel Cole, Hamid Mehran, and Michael Suher — Presented at Fed Reserve Bank of Cleveland, SFA'23, FMA European'23, IBEFA-WEAI'24 (Working paper: accepted to MFS'24 & European FMA'24)
- Data-Driven Response to Opioid Abuse in Banking, with Dusan Ramljak (Working paper)