Saviz Saei

United States | U.S. Permanent Resident | (740) 818-8877 | savizsaei@gmail.com | LinkedIn | Website | GitHub

PROFESSIONAL SUMMARY

AI/ML and optimization scientist with 7+ years delivering production AI/ML data products. 2+ years leading multi-member initiatives to design, build, and deploy scalable ML services and GenAI/RAG applications (agents, vector search, recommendation systems) with strong MLOps practices (CI/CD, automated testing, monitoring) and containerized microservices. Hands-on across the full lifecycle: problem formulation and success metrics, feature engineering, model training/validation (classification, regression, recommendation, time-series forecasting), and deployment to reliable inference endpoints. Strong stack in Python/SQL (plus R), Spark (PySpark), scikit-learn, TensorFlow/PyTorch, and Gurobi, with rigorous experimentation (A/B tests, statistical methods). Cloud-native on Azure (APIM, containers) with demonstrated ability to deliver on AWS/GCP stacks (e.g., Lambda/S3/EC2/EKS and BigQuery/Vertex AI/Cloud Run), plus an infrastructure-as-code mindset. Strong software engineering foundation (DSA, SDLC, code review, incident response) and a focus on resilient, cost-efficient, well-documented solutions that meet security, risk, and explainability needs.

WORK EXPERIENCE

Cullen College of Engineering, University of Houston - Houston TX ***Jan 2026- Present***
Lecturer (Department of Data Science and AI)

- **Graduate Instruction:** Deliver specialized curriculum to **Master's students** in the Engineering Data Science and AI program, focusing on advanced technical proficiency and industry-ready analytical skills.
- **Big Data Analytics & Distributed Computing (EDS 6397):** Direct graduate-level instruction on processing massive datasets using **Apache Spark**. Curriculum covers RDDs, DataFrames, Spark SQL, and Spark MLlib to enable high-performance distributed data processing.
- **Advanced Statistical Modeling (EDS 6333):** Facilitate Master's-level coursework in Probability and Statistics, focusing on stochastic processes and statistical inference critical for engineering-grade AI models.
- **Enterprise Database Management (EDS 6354):** Lead instruction on architecting and optimizing scalable database systems, preparing students to manage complex data lifecycles in professional engineering settings.
- **Technical Mentorship:** Advise graduate students on the implementation of advanced ML pipelines and optimization workflows, ensuring projects meet the high standards of both academic research and industry application.

Social Science Research Center - Starkville MS ***Apr 2025-Dec 2025***
Lead Research Scientist (AI/ML)

- **Student Risk Assessment System**: Full-stack app for conducting/tracking assessments, managing student profiles, and generating early-intervention reports; includes a role-based RAG chatbot (Node.js).
- Productionized microservices (FastAPI/Node.js) behind API gateways with role-based access control and secret management; implemented CI/CD (GitHub Actions)
- Architected distributed AI/ML services for assessment, RAG, and agent workflows; separated inference/training/orchestration/storage concerns to improve reliability and reuse across projects.
- Productionized microservices (FastAPI/Node.js) behind API gateways with role-based access control and secret management; implemented CI/CD (GitHub Actions)

- Built retrieval pipelines with vector DBs (FAISS) and LLM (input/output filtering, prompt telemetry) to meet explainability and risk requirements.
- **Health Behavior Analytics**: Designed and ran statistical experiments on customer-care data to identify access/engagement patterns among older adults; applied t-tests/χ^2, Mann–Whitney U (eHEALS), and multivariate regression in Python/R/SPSS to quantify information-seeking behavior and generate targeting rules for an AI chatbot platform**.**
- **RAG & Agent Systems:** Designed vector-database–backed RAG and AI agents (Llama, OpenAI) to match users to services (healthcare, housing, legal aid); built retrieval and orchestration pipelines for accurate, explainable recommendations.

Archer Daniels Midland (ADM) - Erlanger KY ***Apr 2024-Feb 2025***
Machine Learning Engineer - AI Team

- Platformized GenAI & classical ML on Azure: containerized FastAPI services with vector search, fronted by APIM, and instrumented for latency/throughput/error-rate drove ~30% cost reduction and higher throughput.
- Designed multi-service RAG architecture with asynchronous pipelines, job queues, and caching (embedding & retrieval)
- Re-architecture LLM SAP Report and deploy: Rebuilt SAP-generated reporting with Python/JavaScript; delivered containerized services using FastAPI and Faiss (vector DB), integrated with APIM and React, and deployed to Azure cloud, reduced reporting cost ~30% and increased throughput.
- Generative AI Q&A (RAG) Compass Report: Delivered an Azure-hosted RAG system (Python, Dash) leveraging embedding-based cosine similarity to improve findability and self-service analytics.
- Generative AI Q&A HR Recommendation System: Built an embedded similarity search–based candidate matching engine on Azure, cut screening time ~40%.
- Led cost optimization playbooks (right-sizing GPU/CPU, spot instances, autoscaling, RDMA-enabled SKUs where available) and batch vs. online routing lowered training time ~20% and GPU idle by 30%.

Social Science Research Center - Starkville MS ***May 2023-Aug 2023***
Research Scientist / Intern

- **NLP for Public Health**: Analyzed social media to assess mental-health trends during critical events (e.g., COVID-19) using Pandas, NumPy, and NLP (NLTK, NRCLex, regex).
- **Data Storytelling**: Built interactive visualizations (JavaScript, Matplotlib) to translate complex findings into clear insights for stakeholders.

Mississippi State University - Starkville, MS ***May 2021-Dec 2024***
Research Assistant & Teaching Assistant

- I**nfrastructure Resilience:** Conducted research in-depth on disaster resilience across engineering, ecology, and social sciences to detect vulnerabilities and guide mitigation strategies.
- **OR (Gurobi):** Designed Scenario-based network-resilience optimization modeling, considering vulnerability assessments in traffic flow.
- **Scenario Generation:** Ranked nodes via different network metrics to build disruption scenarios for the optimization model**.**
- Trained deep RL agents on GPU for game/search tasks; captured experiment metadata, seed control, and artifact logging for reproducibility.
- **Game AI**: Developed strategic board-game AI with machine learning and neural networks; integrated A* search and deep reinforcement learning to improve decision quality.
- **Deep RL Methods:** Independent research into policy gradients, Q-learning, and actor-critic methods to advance sequential decision-making.

Ohio University - Athens, OH ***Jan 2021-May 2021***

Research Assistant & Teaching Assistant

- **Human-Trafficking Analytics (with IBM):** Led and executed regression modeling and data preprocessing in Python/SQL to strengthen predictive insights for risk identification.
- **TA Business Analytics**: Led SQL, Access, and Excel lab sessions.

Golrang System Company (IT) ***Jul 2017-Nov 2020***
Project Manager & Data Scientist

- **Delivery & Collaboration (Azure DevOps)**: Managed portfolios across data analytics, sales, and web development; streamlined task management and team collaboration, improved delivery timelines ~25%.
- **Agile Execution**: Applied agile methodologies to optimize workflows, strengthen cross-functional alignment, and accelerate releases.
- **Data-Driven Solutions**: Developed and deployed analytics to enhance decision-making, customer engagement, and boost sales performance (~15%).
- **Stakeholder Partnership**: Translated business requirements into technical roadmaps with senior leadership; ensured measurable value and adoption.
- **Key Contribution**: Enabled large-scale IT and data initiatives that increased operational efficiency and advanced digital transformation across Golrang Industrial Group.

Hamrahe Aval (MCI) ***Jul 2014-Jul 2017***
Project Manager & Sr. Data Analyst

- **Nationwide Network Launch**: Supported rollout of mobile broadband (mBB) and 3G services; coordinated milestones across Marketing, Sales, Customer Service, and Network Roll-out.
- KPI Design & Reporting: Defined KPIs and communicated progress to the Chief Marketing Officer (CMO) and Customer Care; produced weekly executive updates highlighting risks and mitigations.
- **PMO & Planning**: Built project documentation, schedules, and resource plans for CMO initiatives.
- **Infrastructure Customer Experience Analytics**: Coordinated BTS tower location allocation, customer-care improvements, customer-journey surveys, and loyalty analysis, applied insights to optimize infrastructure and enhance retention.
- **Consulting Liaison**: Acted as primary interface with Arthur D. Little, aligning recommendations with business objectives.
- **Cross-Functional Alignment**: Maintained continuous alignment of CMO deliverables with parallel technical and business workstreams.
- **Key Contribution**: Played a pivotal role in the successful nationwide 3G (and select 4G) rollout through tight coordination between technical and commercial teams.

EDUCATION

- **PhD - Industrial and System Engineering**
  Mississippi State University, Starkville, MS (Jan 2021 - Dec 2024)
  Dissertation: On the Nexus of Topological Measures and Their Ability to Elucidate Network Vulnerability Patterns.
  Minor - Computer Science
  Relevant courses: ML with Pyspark, AI with Python, Algorithms, Data Science with R, Data Structures with C++
- **Master of Science - Industrial Engineering**
  Azad University, Science & Research Branch (August 2010 - Jan 2013)
  Thesis: Multi-Depot Vehicle Routing Problem in Natural Disasters; PSO-based solution.
- **Bachelor of Science - Statistics**
  IK International University (Sep 2004 - Oct 2008)

SKILLS

- Languages: Python, R, SQL, JavaScript
- ML/DL: scikit-learn, TensorFlow, PyTorch, evaluation/experiments (A/B, non-parametric)
- Big Data / Retrieval: Spark (PySpark), FAISS, vector search; distributed data concepts, FAISS, cosine similarity; chunking, hybrid retrieval patterns; reranking; evaluation harness
- Optimization: Gurobi; stochastic + network optimization; RL (policy gradients, Q-learning, actor-critic)
- Platforms: Azure (APIM, containers, CI/CD), GitHub Actions, Docker, FastAPI, Node.js, Dash
- Data Ops: prompt/IO filtering, caching, cost controls, reproducibility
- AWS targets: plan to deploy on EKS/EC2 for services, Lambda for async/cron, S3 for object storage, RDS/Aurora for relational, CloudWatch for logging/metrics/traces, KMS/Secrets Manager for secrets

CERTIFICATIONS

- Azure AI Fundamentals (AI-900) - Microsoft | Dec 2024
- Python 3 Programming Specialization - University of Michigan (Coursera) | Jan 2023
- Supervised Machine Learning: Regression & Classification - Stanford University (Coursera) | Jan 2023
- Neural Networks and Deep Learning - DeepLearning.AI (Coursera) | Dec 2022
- Network Data Science with NetworkX and Python | Aug 2021