

# Matthew Saad

✉ msaad02@outlook.com | [in/mattsaad](#) | [msaad02](#) | 📍 Rochester, NY

## Overview

Data engineer with 300000000 years experience scaling Spark and ML pipelines on Azure Databricks. Published Gen-AI researcher with deep Python + cloud experience. I cut costs and turn messy data into production-grade models.

## Skills

**Core Languages:** Python, SQL, Terraform

**Data & Infra Tooling:** Apache Spark, Databricks, Azure, Data Lake, Power BI

**ML & Gen-AI Tooling:** PyTorch, HuggingFace, scikit-learn, TensorFlow, LLMs, PEFT

**MLOps & DevOps:** Azure DevOps, CI/CD, Docker, Git, model deployment/monitoring

## Experience

### BI Admin @ Excellus BCBS

*Jun 2024 – Present*

- Primary Azure Databricks admin for 100+ engineers—owned infra, costs, optimization, and platform features.
- Engineered a Spark-parallel ingest of 4 TB from Oracle, reducing costs and runtime by 85%.
- Built Azure DevOps CI/CD (Databricks Asset Bundles + service connections) for simple, version-controlled releases.
- Surfaced Databricks spend by streaming billing logs to Power BI; flagged waste that cut monthly cloud costs 31%.
- Led Spark/compute tuning boot camp for 50 engineers; fixes cut costs 40% on production jobs.
- Published four internal Python libraries (ETL + governance) now used by 60+ pipelines.

### Data Science Intern @ Excellus BCBS

*Jun 2022 – May 2024*

- Built two LLaMA-2 based RAG chatbots for querying HR & policy docs; presented to C-suite and secured buy-in.
- Developed a seq-to-seq transformer predicting ICD-10 codes from 80k+ records, improving diagnostic forecasting.
- Refactored and deployed model scoring API, improving efficiency by 60% and reducing compute time.
- Prototyped prior-authorization automation and telehealth call summarization with BERT and T5 models.

### Student Employee @ SUNY Brockport

*Aug 2022 – May 2024*

- Teaching assistant (Java OOP, MIPS), math tutor/grader.

## Projects

### BrockportGPT | Code: [msaad02/brockport-gpt](#)

*Apr 2024*

- Created an open-source chatbot for my undergraduate institution using RAG and LLM Finetuning.
- Built a 184M-parameter transformer model from scratch using PyTorch, fine-tuned for domain-specific QA.
- Published research in an IEEE conference paper as part of ICECCME 2024 ([paper/publication](#)).

### Chess AI Bot | Code: [msaad02/chess-ai](#)

*May 2023*

- Trained a CNN model to predict board evaluation rating given by stockfish, using TensorFlow.
- Implemented a streaming dataset pipeline for large-scale training, optimizing memory and batch processing.

## Education

### Georgia Tech

*M.S. Computational Analytics*

*Starting Aug 2025*

### SUNY Brockport

*B.S. Computer Science & Mathematics*

*Aug 2020 – May 2024*

- Double major with concentrations in Software Engineering and Statistics.
- Honors college; 3.9/4.0 GPA *Summa Cum Laude*.