

Muhammad Marzouk Baig

+1-585-301-3526 | baig.muham@northeastern.edu | [linkedin.com/in/muhammadmarzoukbaig](https://www.linkedin.com/in/muhammadmarzoukbaig) | github.com/marzoukbaig14

EDUCATION

Northeastern University – Roux Institute

Master of Science, Artificial Intelligence

- Alfonds Merit Scholarship

Portland, ME

Jan. 2025 – Dec. 2027

University of Rochester

Bachelor of Arts, Computer Science

- Dean's Scholarship Award – Merit Based Academic Scholarship
- President, Chi Phi Fraternity, Kappa Delta Chapter

Rochester, NY

Aug. 2020 – Dec. 2024

Spring 2022 – Fall 2023

EXPERIENCE

1010data (Ten Ten Data)

Software Engineering Intern

New York, NY

Jun. 2022 – Aug. 2022

- Designed and implemented a pip-installable, object-oriented Python library mirroring the Pandas API, converting familiar DataFrame-style method signatures to 1010data's proprietary XML query language via a custom-built translation algorithm; applied OOP principles including operator overloading and method chaining, and validated correctness through an extensive unit test suite.
- Integrated the library end-to-end into the 1010data environment: generated XML was passed from the command line, authenticated via SSO, submitted through custom REST API endpoints built on the internal stack, and executed server-side with results rendered live in the 1010data GUI, enabling a full terminal-to-visualization workflow.
- Built a scripted HTTP request-based web automation layer handling session management and API authentication, replacing a Selenium prototype that failed internal security review; introduced new REST API endpoints that consolidated redundant logic and reduced overall codebase size by over 30%.
- Delivered live technical demos to internal leadership and enterprise clients including Dollar General, executing client-supplied queries on the spot and walking through the full pipeline from library call to server-side execution and GUI rendering.
- Authored dual-audience technical documentation in Confluence: an internal developer guide covering library architecture, OOP design decisions, and the XML translation algorithm, and a client-facing user manual with installation steps, worked examples, and troubleshooting guidance.

Cronus

Full Stack Developer

Rochester, NY

Feb. 2021 – Aug. 2021

- Integrated the Google Places API into a React Native mobile application, implementing real-time address auto-fill with API key management, request throttling, and JSON response parsing to deliver clean, structured location data to UI components.
- Built a two-step user registration and authentication flow using React Native components with client-side form validation and local state management, persisting user profile data to Firestore (NoSQL) with real-time read/write operations.
- Designed and implemented 10+ screens across the application including the Vendor Profile screen, using Figma for mockups and React Native for implementation; established reusable component patterns and maintained consistent design system adherence across the codebase.

Simon Business School, University of Rochester

IT Support Consultant

Rochester, NY

Jul. 2023 – Dec. 2024

- Resolved 100+ hardware and software tickets monthly maintaining 99% system uptime; used Jira for ticket tracking, triage, and workflow management across the support team.
- Developed internal application widgets in Python to automate network-based administrative workflows, using CI/CD pipelines for versioned deployment and applying secure configuration management to protect environment variables and credentials.
- Rebuilt and digitized the IT documentation platform, migrating institutional knowledge into an organized, searchable system and reducing new employee onboarding and training time by 30%.

PROJECTS

MKP Genetic Algorithm – Generating Function Initialization | *Python, PyTorch, NumPy, Jupyter* May 2025 – Present

- Developing a novel initialization framework for genetic algorithms applied to the Multidimensional Knapsack Problem, using combinatorial generating functions to derive per-item feasibility sampling probabilities and bias initial populations toward feasible regions, directly addressing a core bottleneck in evolutionary computation on tightly constrained optimization problems.
- Implemented GPU-accelerated Monte Carlo estimation using PyTorch (10M samples in 5.8s on a T4 GPU); achieved 97.5% feasible initialization vs. 0% under standard random initialization, with 46.5% faster convergence to 95% solution quality, validated across 9 experiments, 3 difficulty levels, and 15 independent runs each.

- Confirmed novelty through personal correspondence with the closest prior work author (Hill, 1999); full codebase, experiments, and results publicly documented on GitHub with daily commits.

GPT-Zero – Transformer from Scratch | *Python, PyTorch*

Spring 2025

- Built a ~10M parameter GPT-style transformer from scratch in PyTorch — implementing scaled dot-product self-attention, multi-head attention, transformer blocks with residual connections and layer normalization, positional embeddings, and autoregressive character-level text generation; no external abstractions, every component hand-implemented from the underlying math.
- Trained on the full Shakespeare corpus (256-token context, 384-dim embeddings, 6 attention heads, 6 transformer layers, 0.2 dropout); preceded by a full progressive curriculum through bigram, MLP, and WaveNet-style language models in a companion repo (nanoLM), culminating in this full transformer implementation.

Listening to Southern Maine | *Python, Pandas, spaCy, Hugging Face Transformers, Scikit-Learn*

Spring 2026

- Built the data cleaning and NLP pipeline for a mixed-methods analysis of 1,702 community survey responses for United Way of Southern Maine across 8 heterogeneous survey instruments; harmonized a sparse ragged schema (8 instruments × 18 question slots) into a unified analyzable dataset, treating missingness as a structural property of each instrument rather than nonresponse.
- Implemented a two-stage free-text classification pipeline — spaCy NER for geographic entity detection followed by zero-shot classification using Facebook’s BART-large-MNLI against 11 community identity categories — achieving an 89.5% final classification rate; cross-validated findings using chi-square tests (Cramér’s V), logistic regression, and PCA to surface structural differences in community needs across economic strata.

Distributed Version Control System | *Rust*

Fall 2023

- Collaboratively designed and implemented a distributed VCS in Rust from the ground up, applying modular architecture and separation of concerns across independently developed components for object storage, commit history, branching, merging, and conflict detection.
- Implemented conflict resolution algorithms and managed distributed state consistency across nodes; the project reinforced applied understanding of distributed systems design, memory-safe systems programming, and collaborative software development workflows in Rust.

Gentrification and University Proximity Research Study | *Python, Scikit-Learn, Pandas, Seaborn* *Fall 2023 – Fall 2024*

- Designed and executed an independent quantitative study applying logistic and linear regression in Scikit-Learn to analyze how proximity to private universities in Monroe County, NY correlates with neighborhood-level gentrification indicators across Rochester, using public census and geospatial data.
- Ran the full supervised ML pipeline: data acquisition, cleaning and wrangling in Pandas, feature engineering of demographic and geospatial variables, cross-validated model training and evaluation, and final results visualization with Matplotlib and Seaborn.

CSP Solver Engine | *Java*

Fall 2022

- Designed and implemented a general-purpose constraint satisfaction engine in Java using backtracking search, constraint propagation, and arc consistency (AC-3); demonstrated the engine on classical CSPs including Australia map coloring and job scheduling problems.

Personal Developer Portfolio | *Next.js, React, Tailwind CSS, JavaScript*

Summer 2025

- Designed and built a responsive personal portfolio web application using Next.js and React with Tailwind CSS, applying component-based architecture, server-side rendering, and responsive UI design; deployed on Vercel.

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, TypeScript, R, C, C++, C#, Swift, SQL, Rust, HTML, CSS

ML / AI: PyTorch, TensorFlow, Scikit-Learn, Hugging Face Transformers, LangChain, NLTK, VADER, Pandas, NumPy, Matplotlib, Seaborn, WordCloud

ML Concepts: Supervised and Unsupervised Learning, Neural Networks, NLP, Sentiment Analysis, Topic Modeling, Feature Engineering, Model Training and Evaluation, Genetic Algorithms, Combinatorial Optimization, Monte Carlo Simulation, GPU-Accelerated Computing

Full Stack / Software Eng.: React, React Native, Next.js, Node.js, Django, .NET, REST API Design, Microservices, OOP, Unit Testing, Object-Oriented Library Design, Modular Architecture, Figma, Firestore, MongoDB

Cloud / DevOps: AWS, Azure, GCP, Docker, Kubernetes, CI/CD Pipelines, Git, GitHub, SVN, Linux, Bash, SSO, OAuth

Data: ETL Pipelines, Data Cleaning, EDA, Statistical Analysis, Data Visualization, Apache Airflow, SQL

Tools: Jupyter, Google Colab, Jira, Confluence, Xcode, VS Code, Postman, Excel, LaTeX

Soft Skills: Independent Research, Technical Documentation, Cross-functional Collaboration, Client Communication, Stakeholder Presentation, Project Management, Agile, Team Leadership, Mentorship, Attention to Detail