Qianyu Zheng

Personal Information

848 Spring St. NW | Portfolio: qzheng75.github.io | zhengqianyu0227@gmail.com
Atlanta, GA, 30308 | LinkedIn: zheng-qianyu-365b60250/ | +1 404-960-4914

Education

Bachelor of Computer Science, **Georgia Institute of Technology**, Atlanta, GA — *August 2022 – May 2025* | GPA 4.0/4.0
Master of Computer Science, **Georgia Institute of Technology**, Atlanta, GA — *January 2026 – May 2027 (expected)*

- Related coursework: AI/ML, algorithms, statistics, linear algebra, computer architecture, computational theory

Skills

- Proficient in **Python (Numpy, Pandas, Matplotlib, scipy, etc.), Java, Database (SQL, MongoDB), Microsoft Excel, Cloud Computing (AWS, Google Cloud), High Performance Computing, AI (PyTorch, NLP, GNN), Linux (Bash), Git.**

Awards/Certifications

- Cloud computing: AWS Certified Cloud Practitioner (CLF-C02) and Machine Learning Specialty (MLS-C01)
- Excel: Microsoft Excel Associate (MO-200) and Microsoft Excel Expert (MO-201)

Experiences

Data Analyst, **Fraunhofer Institute for Wind Energy Systems**, *Bremen, Germany* — *July 2025 – Present (December 2025)*

- Implemented **anomaly detection pipeline** using **Gaussian Mixture Models** to clean terabytes of turbine operational data.
- Built automated **time-series** processing pipelines for aggregating multiple data sources with complex filtering conditions.
- Scaled **big data analytics** using **PySpark** across distributed computing environments to process 6+ months of radar data.

Lecturer, **Python/AI bootcamp at University of Maryland**, *remote* — *July 2024 – August 2024*

- Gave a 4-hour **deep learning** lecture for **200+** AI beginners on **PyTorch** and neural network fundamentals.
- Contributed to course design and teaching methodology as part of a 20-instructor team to achieve **AI education effectiveness**.

Researcher, **Leibniz Institute of Plant Biochemistry**, *Halle (Saale), Germany* — *May 2024 – August 2024*

- Designed scalable **data analysis algorithms** for protein families containing **5+ million** sequences using protein Large Language Models, clustering and network analysis algorithms.
- Researched on biologically-informed data splitting strategies with **evolutionary algorithms** to reduce ML models' validation-test gap on molecular property classification tasks.

Researcher, **Georgia Institute of Technology**, *Atlanta, GA, USA* — *May 2023 - Present*

- Researching the applications of Graph Neural Networks (GNNs) in material science.
- Conduct research to design more **stable machine learning force fields** for molecular dynamics (MD) simulation.
- Obtained experience in High Performance Computing, GNNs, PyTorch, deep learning, and research methodologies.

Teaching assistant, **Georgia Institute of Technology**, *Atlanta, GA, USA* — *January 2023 – May 2024*

- TAed for Introduction to **Object Oriented Programming** course within an instruction team of 42 TAs.
- Mentored 400+ students in **OOP** as **Q&A forum lead** among 42 TAs to provide programming assistance, achieving **4.9/5** effectiveness rating in end-of-semester student survey for four semesters.

Projects

Project Leader, **Natural Language Query for Large Protein Databases** — *August 2024 – January 2025*

- Designed a **multimodal** tool for flexible queries for human protein sequences in UniProt database.
- Leveraged **LLM** Llama 3.1 to generate text queries as training data, a **CLIP** model (**BERT** + ESM) in **contrastive learning** of protein sequence and user query embeddings.
- Developed a Flask application (Flask, HTML/CSS) deployed with **AWS Fargate, ECR, ECS**.
- Obtained experience in multimodal learning, LLM training, Cloud Computing, PyTorch, and deep learning.

Project Developer, **Georgia Institute of Technology Data Science club** — *August 2023 – December 2023*

- Participate in the Workout Of the Day (WOD) prediction project group.
- Use Python to perform **data cleaning** and **feature engineering** pipelines for the downstream machine learning tasks.
- Leverage modern optimization libraries to design an automated **hyperparameter search** pipeline for modeling, improving ML models' performance by 6% in WOD prediction.