Qianyu Zheng

Personal Information

848 Spring St. NW
Atlanta, GA, 30308

Portfolio: qzheng75.github.io
LinkedIn: zheng-qianyu-365b60250/

zhengqianyu0227@gmail.com
+1 404-960-4914

Education

Bachelor of Computer Science, **Georgia Institute of Technology**, Atlanta, GA — *May 2026 (Expected Graduation)*

- **Related coursework**: Machine Learning, Deep Learning, Statistics, Data Structure, Design and Analysis of Algorithms, Graph Theory, Linear Algebra, Combinatorics. — *GPA: 4.0/4.0*

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

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

Experiences

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

- Gave a 4-hour **deep learning** lecture for **200** AI beginners, focusing on **PyTorch** applications in regression tasks.
- 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*

- Interned as a **proteomics** researcher in the computational modeling group in computational chemistry department.
- Design algorithm to perform **exploratory data analysis** on large protein families that are scalable to **five million** sequences.
- Crafted biologically significant data splitting strategies with **clustering** and **evolutionary algorithms**, encouraging fair model evaluation and ensuring the reliability of research outcomes.

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

- Researching the applications of Graph Neural Networks (GNNs) in material science.
- Conduct independent research to design more **stable machine learning force fields** for molecular dynamics (MD) simulation.
- Improve robustness of Machine Learning models for empowering MD simulations for crystal structures, demonstrated by elimination of unstable simulations with limited training data.
- 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.
- Work as the **forum lead** that monitors the Q&A forum. Instructional and **communicational excellency** recognized by students with an average rating of **4.9/5** on overall effectiveness in CIOS surveys.

Projects

Project Leader, **Natural Language Query for Large Protein Databases** — *August 2024 - Present*

- 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**. Now live at nl2prot.org.
- Obtained experience in multimodal learning, LLM training, Cloud Computing, PyTorch, and deep learning.

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

- 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.

Team Leader, **Major-League Hacking HackGT 2023** — *October 2023*

- Led a 4-member team in the building "Plot Visualizer" tool to improve accessibility for neurodivergent STEM students.
- Designed an end-to-end **deep learning** pipeline with image classifier, **YOLO** object detection, and OCR to identify graphs, extract data, and generate visualized data series from complex scientific plots.