# Qianyu Zheng

Computer Science

## PROFESSIONAL SUMMARY

As a knowledgeable and skilled Computer Science professional, I have been dedicated to solving complex problems, developing innovative software solutions, and driving technological advancements. I possess a strong foundation in data science, machine learning, cloud computing, and a commitment to delivering high-quality results.

## PERSONAL DETAILS

📱 +1 404-960-4914

✉️ zhengqianyu0227@gmail.com

📍 848 Spring St NW, Atlanta, 30308, United States

## SKILLS

Machine Learning

Python

Algorithms and Data Structures

Database Management

Statistics

Cloud Computing

Excel

MLOps

ETL

## LANGUAGES

**English**
Highly proficient

**Chinese**
Native speaker

**German**
B2

## WORK EXPERIENCE

**Jul, 2025 - Dec, 2025**

### Data Analyst

*Fraunhofer Institute for Wind Energy Systems - Bremen, Germany*

- Implemented anomaly detection pipeline with **Pandas/Scikit-learn** 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 & SQL** to process 6+ months of remote sensing data (radar, lidar, etc.).
- Performed geographical data visualization with **Matplotlib, rasterio,** and **cartopy.**

**Jul, 2024 - Aug, 2024**

### Lecturer

*Python/AI "Scotch" Bootcamp, University of Maryland - Remote*

- 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**.

**May, 2024 - Jul, 2024**

### Computational Scientist

*Leibniz Institute of Plant Biochemistry - Halle (Saale), Germany*

- Designed scalable data analysis algorithms for protein families containing **5+ million** sequences using protein Large Language Models, **clustering** and network analysis algorithms.
- Researched on chemically significant data splitting strategy for molecular machine learning tasks and implemented an innovative approach with **evolutionary algorithms**.

**Jun, 2023 - Present**

### Computational Scientist

*Georgia Institute of Technology - Atlanta*

## CERTIFICATES

### Certified Cloud Practitioner (CLF-C02)

*Amazon Web Services |*
Foundational, high-level understanding of AWS Cloud, services, and terminology.

### Machine Learning Specialty (MLS-C01)

*Amazon Web Services |*
Expertise in building and deploying machine learning solutions in the AWS Cloud.

### Microsoft Excel Expert (MO-201)

*Microsoft |*
Expertise in Excel 2019, such as creating, managing, and distributing professional spreadsheets for a variety of specialized purposes and situations.

---

- Conducted research to improve reliability of Machine Learning Interatomic Potentials for molecular dynamics (MD) simulation with High Performance Computing, **GNNs**, **PyTorch**, **deep learning**, and research methodologies.

**Jan, 2023 - Jun, 2024**

### Teaching Assistant

*Georgia Institute of Technology - Atlanta*

- TAed for Introduction to **Object Oriented Programming** course within a 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.

## EDUCATION

### Bachelor of Science in Computer Science

*Georgia Institute of Technology - Atlanta*
**Aug, 2022 - Jun, 2025**

### Master of Science in Computer Science

*Georgia Institute of Technology - Atlanta*
**Aug, 2025 - Present**

## PROJECTS

**Aug, 2024 - Jan, 2025**

### Natural Language Query for Large Protein Databases

*Atlanta*

- Designed a **multimodal** tool for text queries for protein sequences in UniProt database.
- Leveraged **LLM** Llama 3.1 with **Langchain** to generate text queries as training data and PyTorch to train a **CLIP** model (**BERT** + ESM) with **contrastive learning** of protein sequence and user query embeddings. Benchmark against NER-based method with SpaCy shows a **2x accuracy** of the designed method.
- Developed a Flask application (Flask, HTML/CSS) deployed with **Docker**, **AWS Fargate, Lambda, ECR, ECS** and **S3**.

**Aug, 2024 - Dec, 2024**

### Workout Of the Day (WOD) prediction with Data Science @ Georgia Tech

*Atlanta*

- Designed and implemented **data cleaning** and **feature engineering** pipelines for the downstream machine learning tasks.
- Leveraged modern optimization libraries to design an automated **hyperparameter search** pipeline for modeling, improving ML models' performance by 6% in WOD prediction.