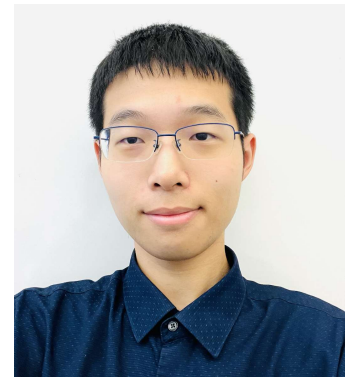# Qianyu (Kenneth) Zheng

Adress: Weizemkampstraße 133, 28199 Bremen
Telephone: +1 4049604914
Email: zhengqianyu0227@gmail.com
Linkedin: www.linkedin.com/in/zheng-qianyu-365b60250
Website: qzheng75.github.io

Motivation

I am a high-achieving master's student at Georgia Institute of Technology with a perfect GPA and extensive hands-on experience applying machine learning and data analytics to solve complex real-world problems across research institutions in the US and Europe. My expertise spans deep learning, cloud computing, data analysis and big data processing. I'm actively looking for opportunities to leverage my expertise in bridging the gap between computing and scientific discoveries.

Work Experiences

07/2025 – now

Data Analyst at Fraunhofer Institute for Wind Energy Systems in Bremen
- 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.**

07/2024 – 08/2024

Dozent für Python/AI-„Scotch"-Bootcamp an der University of Maryland
- Gave a 4-hour **deep learning** lecture for **200+** AI beginners on **PyTorch** and neural network fundamentals and contributed to course design and teaching methodology as part of a 20-instructor team to achieve **AI education effectiveness**.

05/2024 – 08/2024

Computational Scientist at Leibniz Institute of Plant Biochemistry in Halle (Saale)
- Designed scalable data analysis algorithms for protein families containing **5+ million** sequences using protein Large Language Models, **clustering** and network analysis algorithms.
- Participated in protein binding possibility estimation research. Implemented big data handling pipelines with duckdb and **PySpark** and deep learning solutions featuring **transformer** architecture with **PyTorch.**

05/2023 – now

Computational Scientist at Georgia Institute of Technology in Atlanta, USA
- 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.

Teaching Assistant am Georgia Institute of Technology in Atlanta, USA

- 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.

Ausbildung

2022 – 2025 | GPA 4.0/4.0

Bachelor of Science in Informatik, **Georgia Institute of Technology**, Atlanta, USA

2025 – 2027 | GPA 4.0/4.0

Master of Science in Informatik, **Georgia Institute of Technology**, Atlanta, USA

Projekte

08/2024 – 01/2025

Natural Language Query für große Proteindatenbanken

- 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**.
- Obtained experience in multimodal learning, LLM training, Cloud Computing, PyTorch, and deep learning.

08/2023 – 12/2023

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

- 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.

Language

- English – equivalent to native level
- German – planning to take the B2 exam in December 2025

Certificates

Amazon Web Services: Certified Cloud Practitioner (CLF-C02)

- Foundational, high-level understanding of AWS Cloud, services, and terminology.

Amazon Web Services: Machine Learning Specialty (MLS-C01)

- Expertise in building and deploying machine learning solutions in the AWS Cloud.

Excel: Microsoft Excel Expert (MO-201)

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