

# Qianyu (Kenneth) Zheng

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



## Kurzprofil

Ich bin ein leistungsstarker Masterstudent am Georgia Institute of Technology mit umfangreicher praktischer Erfahrung in der Anwendung von maschinellem Lernen und Datenanalyse zur Lösung komplexer realer Probleme an Forschungseinrichtungen in den USA und Europa. Meine Expertise umfasst Deep Learning, Cloud Computing, Datenanalyse und Big-Data-Verarbeitung. Ich suche aktiv nach Möglichkeiten, mein Fachwissen einzusetzen, um die Lücke zwischen Computing und wissenschaftlichen Entdeckungen zu schließen.

## Berufliche Erfahrungen

07/2025 – heute

### Data Analyst bei Fraunhofer-Institut für Windenergiesysteme in Bremen

- Implementierung einer Anomalie-Erkennungs-Pipeline mit **Pandas/Scikit-learn** zur Bereinigung von **Terabytes** an Betriebsdaten von Windkraftanlagen.
- Entwicklung automatisierter **Zeitreihen-Verarbeitungs-Pipelines** zur Aggregation mehrerer Datenquellen mit komplexen Filterbedingungen.
- Skalierung von **Big-Data-Analysen** mittels **PySpark & SQL** zur Verarbeitung von über 6 Monaten Fernerkundungsdaten (Radar, Lidar, etc.).
- Durchführung geografischer Datenvisualisierung mit **Matplotlib, rasterio** und **cartopy**.

07/2024 – 08/2024

### Dozent für Python/AI-„Scotch“-Bootcamp an der University of Maryland

- Durchführung einer 4-stündigen **Deep-Learning**-Vorlesung für über **200 AI-Anfänger** zu PyTorch und den Grundlagen neuronaler Netze sowie Mitwirkung in einem Team von 20 Dozenten zur Verbesserung der Effektivität der AI-Ausbildung.

05/2024 – 08/2024

### Computational Scientist am Leibniz-Institut für Pflanzenbiochemie in Halle (Saale)

- Entwicklung skalierbarer Datenanalyse-Algorithmen für Proteinfamilien mit **über 5 Millionen** Sequenzen unter Verwendung von Protein-Large-Language-Models, Clustering- und Netzwerkanalyse-Algorithmen.
- Mitwirkung an der Forschung zur Abschätzung von Proteinbindungsmöglichkeiten. Implementierung von Big-Data-Verarbeitungs-Pipelines mit **DuckDB** und **PySpark** sowie Deep-Learning-Lösungen mit **Transformer**-Architektur in **PyTorch**.

05/2023 – heute

### Computational Scientist am Georgia Institute of Technology in Atlanta, USA

- Durchführung von Forschungsarbeiten zur Verbesserung der Zuverlässigkeit von Machine-Learning-Interatomic-Potential für Molekulardynamik-Simulationen (MD) mittels High-Performance-Computing, GNNs, PyTorch, und Deep Learning.

01/2023 – 05/2024

## Teaching Assistant at Georgia Institute of Technology in Atlanta, USA

- Tätigkeit als Teaching Assistant für den Kurs „**Introduction to Object Oriented Programming**“ in einem Team von 42 TAs.
- Betreuung von **über 400 Studierenden** in objektorientierter Programmierung als Leiter des Q&A-Forums zur Bereitstellung von Programmierunterstützung, mit einer Bewertung von **4,9/5 für die Lehreffektivität** in der Befragung über vier Semester hinweg.

## Education

---

2022 – 2025 | GPA 4.0/4.0

Bachelor of Computer Science, **Georgia Institute of Technology**, Atlanta, USA

2025 – 2027 | GPA 4.0/4.0

Master of Computer Science, **Georgia Institute of Technology**, Atlanta, USA

## Projects

---

08/2024 – 01/2025

### Natural Language Query for Large Protein Databases

- Entwicklung eines **multimodalen** Tools zur Textabfragen für UniProt-Datenbank.
- Einsatz von LLM Llama 3.1 mit **Langchain** zur Generierung von Textabfragen als Trainingsdaten sowie PyTorch zum Trainieren eines **CLIP-Modells** (BERT + ESM) von Proteinsequenz- und Nutzerabfrage-Embeddings. Benchmark-Vergleich mit NER-basierter Methode mittels SpaCy zeigt eine **doppelt** so hohe **Genauigkeit** der entwickelten Methode.
- Entwicklung einer Flask-Anwendung (**Flask, HTML/CSS**), die mit **Docker, AWS Fargate, Lambda, ECR, ECS** und **S3** bereitgestellt wurde. Die Anwendung ist jetzt online unter: [www.nl2prot.org](http://www.nl2prot.org)
- Erwerb von Erfahrung in Deep Learning, LLM-Finetuning, Cloud Computing und PyTorch.

08/2023 – 12/2023

### Workout Of the Day (WOD)-Vorhersage mit Data Science @ Georgia Tech

- Entwurf und Implementierung von **Datenbereinigung-** und **Feature-Engineering-Pipelines** für nachgelagerte Machine-Learning-Aufgaben.
- Einsatz moderner Optimierungsbibliotheken zur Entwicklung einer automatisierten **Hyperparameter-Search**-Pipeline für die Modellierung, wodurch die Leistung der ML-Modelle bei der WOD-Vorhersage um 6 % verbessert wurde.

## Language

---

- **Englisch** – fließend (entspricht muttersprachlichem Niveau)
- **Deutsch** – fortgeschrittene Kenntnisse, B2-Prüfung geplant für Dezember 2025

## Certificates

---

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

- Grundlegendes, übergeordnetes Verständnis von AWS Cloud, Diensten und Terminologie.

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

- Expertise im Aufbau und der Bereitstellung von Machine-Learning-Lösungen in AWS.

### Excel: Microsoft Excel Expert (MO-201)

- Expertise in Excel 2019, wie z. B. Erstellung, Verwaltung und Verteilung professioneller Tabellenkalkulationen für verschiedene spezialisierte Zwecke und Situationen.