

Qianyu Zheng

Master's Student, Georgia Institute of Technology, Atlanta, GA, USA

zhengqianyu0227@gmail.com — +1 (404) 960-4914 — www.linkedin.com/in/zheng-qianyu-365b60250/ — Personal Website

RESEARCH INTERESTS

AI for Science, Machine Learning, Data Science, Machine Learning Interatomic Potentials, Large Language Model Agents, Inverse Design of Materials, Molecular Dynamics Simulation

EDUCATION

Georgia Institute of Technology, Atlanta, USA

Master of Science in Computer Science

August 2025 – May 2027 (Expected)

Cumulative GPA: 4.00/4.00

Georgia Institute of Technology, Atlanta, USA

Bachelor of Science in Computer Science

August 2022 – May 2025

Cumulative GPA: 4.00/4.00

RESEARCH EXPERIENCE

Georgia Institute of Technology

Undergraduate & Graduate Researcher, Fung Group

Atlanta, GA, USA

May 2023 – Present

- Developed a physics-informed pretraining strategy for machine-learning interatomic potentials (MLIPs) featuring pre-training on augmented low-fidelity datasets labeled with a trained embedded atom method potential to encode short-range repulsion and bonding physics; validated across three chemistries (silica, phosphorus, MPTrj) and three architectures (CGCNN, TorchMD, M3GNet). Experiments on MD simulations demonstrated reduced unphysical MD artifacts (atom overlap, thermal disorder) and better recovery of AIMD trajectories, benchmarked against regularization methods. *Published in Journal of Chemical Information and Modeling (JCIM, first author).*
- Built a differentiable toolbox exposing geometric structure properties (bond angle, symmetry, local coordination environment) as auto-grad-optimizable loss functions, integrated with LLM agents via MCP to enable natural language instruction-driven structure editing; achieved 87.5% success on 40 natural-language modification tasks, outperforming Claude Code and toolbox-free baselines, and designed a multi-agent propose-act-feedback system that achieved 9/10 success tripling target band gaps via inverse design. *Manuscript in preparation; to be posted on arXiv.*
- Leading an agentic descriptor-mining project that processes band structure and DOS DFT outputs into a structured feature bank and applies SISISO within an agent framework to discover interpretable descriptors for superconducting T_c , tested on a dataset of $\sim 8,000$ conventional superconductors; investigating descriptor transferability across doped parent structures and superconductor families. *Ongoing.*

Lila Sciences, Inc.

Intern, Physical Science Department, Simulation Team

Cambridge, MA, USA

May 2026 – August 2026

- Built an agentic framework for automated hyperparameter optimization of machine-learning interatomic potentials and property-prediction models, integrated into internal research infrastructure; models tuned with the framework reached state-of-the-art internal benchmark performance and entered production use.
- Developed a multi-agent auto-research system that autonomously evolves MLIP architectures through mathematical reasoning over model design, formulating new proposals concurrently with in-flight optimization runs for higher experimental throughput than sequential auto-research loops. Demonstrated on recovering strong-baseline performance from an underperforming architecture and on extending a model to a property outside its original scope.

Fraunhofer Institute for Wind Energy Systems

Research Intern, Wind Energy / Wake Simulation

Bremen, Germany

August 2025 – December 2025

- Curated cases for benchmarking wake models by processing one year of TB-scale, multi-instrument wind measurement data (dual-Doppler radar, lidar, disdrometer, met stations) from the AWAKEN project using NetCDF and Spark-based big-data pipelines; invented a data clean methods with the local contrast filter to discard cases with unphysical spikes in radar measured data, identified meteorological/seasonal correlates of artifact occurrence, and validated that flagged cases showed degraded wind speed/angle accuracy against reference instruments.
- Used the cleaned, artifact-free dataset to benchmark two engineering wake models across the full AWAKEN domain (two wind farms, $> 100 \text{ km}^2$), comparing their ability to capture cross-farm wake interaction under varying turbulence regimes. *(JRSE, first author).*

Leibniz Institute of Plant Biochemistry

Research Intern, Machine Learning

Halle (Saale), Germany

May 2024 – August 2024

- Explored large protein families by representing them as sequence-similarity graphs (embedding-distance-weighted edges) and evaluated graph sparsification, community detection, and centrality metrics to identify functionally important sequences on cytochrome c and influenza protein families, assessing robustness of detected communities across sparsification thresholds.
- Designed a cluster-based, distance-maximizing data split strategy for non-i.i.d. protein sequences, combining ESM embedding clustering with a surrogate-model-assisted evolutionary search (random forest fitness, size/label-balance aware) to construct train/val/test partitions; outperformed random split and SpanSeq on classification tasks with cytochrome c and influenza families.

PROJECTS

Query Large Protein Databases with Natural Language

Atlanta, GA, USA

Project Lead

September 2024 – January 2025

- Led a 2-person team to build NL2Prot, a CLIP-style contrastive learning system aligning protein sequence embeddings (ESM2) with natural language description embeddings (BERT) to enable free-text retrieval over professional protein databases (UniProt); curated a training corpus of 10k description-protein pairs by pairing UniProt entries with LLM-generated layman and professional descriptions.
- Achieved up to 94.05% top-50 retrieval accuracy on layman-language queries and 83.95% top-100 accuracy on professional queries, substantially outperforming an NER/keyword-extraction baseline (67.33%/24.55%) while cutting query latency from over 24 hours to under 40 seconds; deployed a quantized inference pipeline via Flask and Docker on AWS (ECS/Fargate, ECR, Route 53) at nl2prot.org.

PUBLICATIONS (pending now)

Journal articles

- [1] Q. Zheng and V. Fung, “Improving Reliability of Machine Learning Interatomic Potentials with Physics-Informed Pre-training,” *Journal of Chemical Information and Modeling*, 2026. doi:10.1021/acs.jcim.6c00826.
- [2] Q. Zheng, L.-Y. (Lilian) Hung, A. Jordan, and J. Gottschall, “An extensive framework for preparing dual-Doppler radar measurements for wake model validation: Application to the AWAKEN large-scale field experiment,” *Journal of Renewable and Sustainable Energy*, vol. 18, no. 4, art. 043302, 2026. doi:10.1063/5.0331761.

AWARDS & CERTIFICATIONS

Graduate Research Assistantship

Atlanta, GA, USA

Funded assistantship supporting full-time research as a Master’s student in the Fung Group

August 2026 – Present

President’s Undergraduate Research Award

Atlanta, GA, USA

Competitive, faculty-mentored research award (PURA) to support undergraduate research

August 2024 – December 2024

Faculty Honors

Atlanta, GA, USA

Awarded each semester for achieving a 4.0 GPA

Every semester in Bachelor studies

AWS Certified Cloud Practitioner

Remote

Foundational certification validating knowledge of AWS Cloud concepts, services, security, and pricing

February 2024

Microsoft Excel Expert

Remote

Advanced Excel skills, including data modeling, complex formulas and visualizations

April 2024

AWS Certified Machine Learning Specialty

Remote

Advanced certification validating ability to design, build, and deploy ML solutions on AWS

March 2024

TEACHING EXPERIENCES

Scientific Computing From Scratch Bootcamp

Remote

Instructor

July 2024 – August 2024

- Served as an instructor for the 2024 iteration of a free, remote summer bootcamp (scotch.wangyq.net) introducing ~200 science-major students to scientific computing and deep learning with Python and PyTorch.
- Led the lecture on deep learning and PyTorch fundamentals (autodiff, building MLPs from scratch) as main instructor, and served as assistant instructor for the lecture on graph neural networks for molecular applications.

GaTech CS 1331: Introduction to Object-Oriented Programming

Atlanta, GA, USA

Teaching Assistant

January 2023 – May 2024

- Taught and supported a large-enrollment (~500 students/semester) introductory course on object-oriented programming principles and JavaFX, working alongside a team of ~25 TAs across four semesters to deliver grading, office hours, and weekly recitations.

- Served as Q&A forum lead for two semesters, owning homework/exam logistics announcements and coordinating TA responses to student inquiries to ensure consistent, timely guidance across the course.

REFERENCES

Prof. Rafael Gómez-Bombarelli

Paul M. Cook Career Development Professor, Associate Professor of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; Co-Founder & Chief Scientific Officer, Physical Sciences, Lila Sciences

E-mail: rafagb@mit.edu

Scholar Profiles: Learning Matter Lab (MIT) — Google Scholar — LinkedIn

Prof. Victor Fung

Assistant Professor, School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

E-mail: victorfung@gatech.edu

Scholar Profiles: Georgia Institute of Technology - Faculty Profile — Google Scholar — LinkedIn

Prof. Julia Gottschall

Chief Scientist & Chief Science Coordinator, Fraunhofer Institute for Wind Energy Systems IWES; Professor, University of Bremen, Bremen, Germany

E-mail: julia.gottschall@iwes.fraunhofer.de

Scholar Profiles: ResearchGate — LinkedIn