CAILONG HUA

(651)675-6613 ◇ hua00023@umn.edu◇ huacailong.github.io ◇ www.linkedin.com/in/cailongh ◇ huacailong

A **Ph.D. candidate** with expertise in LLMs, Machine Learning, Signal Processing and Data Analysis, driven by a passion for pushing the boundaries of AI in the medical area.

EDUCATION

Ph.D. in Electrical Engineering with Computer Science Minor *08/2019 - 12/2025 (Expected)*
University of Minnesota - Twin Cities | GPA: 3.95/4.0

Master of Science in Control Systems *09/2016 - 09/2017*
Imperial College London, London | GPA: 75.4/100 (Distinction)

Bachelor of Science in Automation Engineering *09/2015 - 07/2016*
Politecnico di Milano, Milano | GPA: 110/110

Bachelor of Engineering in Electronic Information and Engineering *09/2012 - 07/2016*
Tongji University, Shanghai | GPA: 89.52/100(4.45/5.0)

RESEARCH EXPERIENCE

Single Molecule Modeling of Muscle Proteins *06/2021 - Present*
Ph.D. Thesis Project
- Designed force spectroscopy experiments to characterize muscle proteins associated with muscular dystrophy
- Automated analysis of experimental data with Matlab, reducing processing time from days to hours
- Resolved electrostatic discharge issues in atomic force microscopy experiments through collaborative troubleshooting
- Conducted statistical analysis of experimental data to model muscle proteins
- Reconstructed energy landscapes of proteins from noisy and uncertain data
- Collaborated with biochemists to investigate the impact of different expression systems on protein behavior

Classification models for single molecule force spectroscopy (SMFS) data *03/2023 - Present*
Ph.D. Thesis Project
- Compiled single-molecule force spectroscopy datasets from six biologically significant proteins
- Developed a physics-based Monte Carlo simulation engine to generate synthetic SMFS data
- Designed a novel deep learning model incorporating protein unfolding physics to classify single-molecule data
- Reduced data analysis time from several days to under an hour using the proposed model

Clustering models for single molecule force spectroscopy data *12/2024 - Present*
Ph.D. Thesis Project
- First to uncover heterogeneous mechanical protein domains using data-driven clustering models
- Proposed a novel physics-aware deep clustering model to capture biologically meaningful relationships
- Successfully identified heterogeneous mechanical domains across multiple biologically significant proteins

Error Quantification for Non-Equilibrium Experiments with Limited Data *06/2020 - 03/2025*
Ph.D. Thesis Project
- Developed an algorithm for quantifying errors in non-equilibrium experiments
- Released a Python-based toolbox for error quantification and validated on simulated spring-mass system
- Validated the algorithm through Optical Tweezers experiments under high-noise conditions

PROFESSIONAL EXPERIENCE

Data Scientist Intern *05/2024 - 08/2024*
University of Phoenix, Inc.
- Fine-tuned large language models (LLMs) on task-specific data
- Evaluated various parameter-efficient fine-tuning (PEFT) methods on different pre-trained LLMs
- Developed a chatbot in Amazon Web Service (AWS) to interface with fine-tuned LLMs

PROJECTS

Determining Causality in Protein Unfolding Pathways *09/2022 - 12/2022*
- Uncovered causal relations from protein pulling data
- Quantified the change of effects among protein properties
- Investigated causality under the assumption of unobserved data

Different Sampling Methods in Energy-Based Models (EBMs) *09/2022 - 12/2022*
- Explored Langevin-based sampling methods in a 2D localization problem
- Applied sampling techniques to EBMs, reducing computational workload and improving performance on the MNIST dataset

Deep Learning Based Pose-Guided Person Image Generation *09/2021 - 12/2021*
- Utilized Generative Adversarial Networks (GANs) to generate realistic images of individuals in target poses

- Enhanced person transfer performance using the HUMBI dataset

Performance Discrepancy in Natural Language Processing (NLP) *09/2020 - 12/2020*
- Explored the impact of fine-tuning layers and pooling strategies on BERT (a transformer based machine learning technique for NLP) performance
- Achieved improvements in accuracy and F1 score over the baseline method

Advanced Control Techniques for Dynamic Systems *10/2016 - 05/2020*
- Implemented PID control on both simulated and real-world unmanned aerial vehicles (UAVs).
- Applied PID and model predictive control (MPC) to stabilize an inverted pendulum system.
- Designed and implemented H_∞ control for the experimental aircraft.
- Modeled and developed linearized control strategies for autonomous grocery trolleys.
- Applied double PID, LQR, H_∞, and μ-synthesis control techniques for robust stabilization.

Automating Game Playing Using Deep Reinforcement Learning *09/2019 - 12/2019*
- Designed an unsupervised agent to autonomously play the Flappy Bird game
- Applied Reinforcement Learning techniques to AI agent to learn the optimal policy and survive indefinitely

PUBLICATIONS

Journal Publications
- **Hua, C.**, Rajaganapathy, S., Slick, R.A., Vavra, J., Muretta, J.M., Ervasti, J.M., and Salapaka, M.V., "A physics-augmented deep learning framework for classifying single molecule force spectroscopy data." *In Forty-second International Conference on Machine Learning.(ICML 2025).*
- **Hua, C.**, Zhang, Y., Singh, V., Muretta, J.M., Ervasti, J.M., and Salapaka, M.V., "Uncovering Heterogeneous Mechanical Protein Domains via Clustering, Including a Novel Physics-Aware Model." In preparation.
- **Hua, C.**, Vavra, J., Powers, J., Muretta, J.M., Ervasti, J.M., and Salapaka, M.V., "Multiple modes of atomic force microscopy reveal distinct mechanical properties for dystrophin and utrophin not manifest by small fragments." Under review, *PNAS*.
- Rajaganapathy, S.*, **Hua, C.*** and Salapaka, M.V., "Quantifying errors in the Jarzynski estimator." Under review, *Physical Review E*. (*Co-first authors*).
- **Hua, C.**, Slick, R.A., Vavra, J., Muretta, J.M., Ervasti, J.M., and Salapaka, M.V., "Two operational modes of atomic force microscopy reveal similar mechanical properties for homologous regions of dystrophin and utrophin." *bioRxiv*, 2024:2024-05. In preparation.
- Ramirez, M.P., Rajaganapathy, S., Hagerty, A.R., **Hua, C.**, Baxter, G.C., Vavra, J., Gordon, W.R., Muretta, J.M., Salapaka, M.V., and Ervasti, J.M., "Phosphorylation alters the mechanical stiffness of a model utrophin fragment." *Journal of Biological Chemistry*, 299(2), 2023.

Talks, Presentations, and Posters
- **Hua, C.** and Salapaka, M.V., "Assessing mechanical properties of a utrophin fragment with two operational modes." *Bulletin of the American Physical Society*, 2024.
- **Hua, C.**, Rajaganapathy, S., and Salapaka, M.V., "Quantifying errors in the Jarzynski estimator." *9th Midwest Workshop on Control and Game Theory*, Poster, 2023.
- Rajaganapathy, S., **Hua, C.**, and Salapaka, M.V., "Confidence bounds for the Jarzynski estimator." *APS March Meeting Abstracts*, Vol. 2022, pp. S09-007, 2022.

Reviewed for contributed articles in
- International Conference on Learning Representations (ICLR)
- American Control Conference (ACC)
- Proceedings of the National Academy of Sciences of the United States of America (PNAS)

SKILLS

Technical: Python, AWS, Matlab, Simulink, Latex, C, C++, R
Machine Learning Tools: Tensorflow, Pytorch, Scikit-Learn, Seaborn, Pandas, etc.
Communication: English; Chinese; Italian

TEACHING EXPERIENCE

Teaching Assistant *09/2020 - Present*
University of Minnesota, Twin Cities
- Assisted in multiple labs: Linear Control Systems Lab, State Space Control Systems Lab, Electric Drives Lab, Digital System Design Lab and Microcontrollers Lab
- Transitioned the entire laboratory to an online format during the COVID time
- Modernized aging experiments using hardware-in-the-loop (HIL) control and developed new instructional manuals
- Mentored new teaching assistants
- Trained new lab members on high-precision instruments, including atomic force microscopy (AFM) and optical tweezers (OT)