Richard Wang

408-431-1437 richardwang007@gmail.com

Education

California Institute of Technology (Caltech)

B.S. Computer Science; Class of 2023

COURSEWORK:

Computer Science: Computing Systems, ML and Data Mining, Programming Methods, Decidability and Tractability, Data Structures, Compilers and Interpreters, Computer Architecture, Expert Systems, Learning Systems, Vision Neuromechanics Theory and Computation, GPU Programming, Algorithms, Information Theory

Math: Discrete Math, Abstract Algebra, Calculus on Manifolds, Differential Equations, Probability and Stats, Number Theory, Applied Linear Algebra, Probability Models, Mathematical Options Pricing

Experience

MACHINE LEARNING RESEARCHER, PERFORMANCE STAR LLC – 2023 - PRESENT

- Developed full ML data pipelines, training, and inference pipelines, servers, and API from scratch using Python, Docker, Kubernetes, MLflow, Tensorflow, Tensorflow Serving, and Flask.
- Led research in generative models for high-dimensional, sparse time-series data from semiconductor fabrication; focused on predictive maintenance and anomaly detection for on-device real time inference.
- Designed and implemented siamese and generative deep learning architectures, using spectral preprocessing techniques for novel learning frameworks.
- Developed hybrid methods (mathematical + deep learning) for modeling chaotic dynamical systems
- Research and development into Dynamic Mode Decomposition (DMD), Complex Valued Neural Networks (CVNN), Wavelet analysis, VAEs, transformers, and other SoA deep learning and mathematical techniques.

DATA ANALYSIS INTERN, BLUESKY – 2023

- Research and development for automated query optimization on Snowflake.
- Implemented machine learning solutions for time series modeling on query metadata.

MACHINE LEARNING RESEARCH INTERN, MIT LINCOLN LABS – 2022

- Research and analysis on various SoA semantic segmentation ML models for government projects.
- Designed and implemented image-to-image models using Pytorch.
- Implemented ML pipeline with Hydra and DVC.

MACHINE LEARNING RESEARCH INTERN, THERMO FISHER SCIENTIFIC – 2021

- Developed machine learning solutions for non-identical image-mask homography estimation and image manipulation AI.
- Implemented virtual machine automation for ML ops deployment.

CALTECH DEANS TUTOR – 2020-2023

- Teaching systems programming, math, and analytical physics.

DATA ANALYSIS INTERN, UNIVERSITY OF SOUTHERN CALIFORNIA (USC) – 2020

- Analyzed single-cell sequencing data, comparing gene expression profiles between various cancer cell lines.

- Performed unsupervised gene cluster analysis to investigate drug resistance mechanisms.

DATA ANALYSIS INTERNSHIP, UNIVERSITY OF MICHIGAN ANN ARBOR – 2018-19

- Analyzed high-throughput compound screening data using statistical methods.
- Discovered a novel compound for cancer therapy and won National Regeneron Scholar Award.

Awards and Publications

- 2024 Publication and patent in preparation: *Siamese Learning and Dynamic Mode Decomposition in Semi-supervised Predictive Maintenance for Recipe-Agnostic Semiconductor Fabrication*
- 2022 National Association of Basketball Coaches (NABC) Academic Excellence Award
- 2021 Presentation at Thermo Fisher Scientific Innovation Day (Poster: *Homography Estimation for Non-Identical Image Mask Pairs*)
- 2019 NuevaHacks NEAR Protocol Best Util (NuevaHacks Prize Winner)
- 2019 Regeneron Scholar (Paper: *Identify Novel Natural Products with Inhibitory Effects on HOXA-9-induced Acute Leukemia through in vitro functional Assays*)
- 2019 Presentation at Pathobiology for Investigators, Students and Academicians Conference
- 2018, 2019 President's Award for Volunteer Work
- 2017 NorCal Sports TV Asian-American All Star (www.youtube.com/watch?v=E67m1DTT-Gk)
- 2016 US Chess Federation High School National Chess Team Champion (https://news.harker.org/harker-students-win-individual-and-team-national-chess-championships/)
- 2015 NorthSouth NorCal Top10 Basketball (https://www.northsouthbasketball.com/college-players)

Volunteer Work and Non-Profit

YOUTH BASKETBALL COACH, HOOPRIGHT – 2024

- Coaching 6th grade boys basketball for AAU in the Bay Area.

GUITAR LESSONS, CANON MUSIC – 2024

- Teaching youth guitar lessons in various genres in the Bay Area.

BASKETBALL COACH, CUPERTINO HOOPS – 2016-19; 2023-24

- Youth coaching 3rd-6th grade with a local basketball rec-league.

Key Skills

Programming: Python, C/C++, PyTorch, Tensorflow, x86-64 assembly, Mathematica, JS, Java, SQL

Math: Functional Analysis, Advanced Algorithms and Discrete Maths, Complexity Theory, Probability and Statistics

Apps/Frameworks: IDA Pro, Wireshark, git, aircrack, MySQL, nodeJS, sqlite, LAMP, MLflow, Kubernetes, Docker, Jupyter

Projects

SELF DRIVING MODEL CAR

Engineered model car with self-driving capabilities using Intel Real-Sense depth camera and ROS. Autonomous navigation, on-device ML. In Python.

COMPILER + INTERPRETER

Compiler for Java as well as an interpreter for Java bytecode (basically a mini JVM). In C.

WEB PROXY

Multi-threaded web proxy in C++. Using a concurrent hash table to implement a LRU cache system.

ML SONAR FISH TRACKING

An Alaskan government project to detect, track, identify, and count fish via machine learning and motion tracking techniques. Deployed real-time ML solutions for in-water sonar data.

MINI OS

Simple monolithic operating system from scratch. File system, kernel threads, different programs. In C.

Personal

Bilingual: English (Native), Mandarin (Native)

Misc: Former Caltech Men's Basketball Team; Former President of Cooking Club; CTF and Hackathon enthusiast; Former Caltech Quant Trading Club, Former Caltech Chess Club, Guitar and Piano nerd

Astounding advancements in artificial intelligence—particularly in limited memory models, embedding techniques, and physics-informed learning—fuel my drive to pursue a PhD in Computer Science at Columbia University. My research interests in multimodal models, dimension reduction, and statistical learning have been shaped by diverse academic and industry experiences, and I am eager to join Columbia's interdisciplinary community to advance the frontier of AI research.

Current Work: At PerformanceStar, where I work as a Machine Learning Research Engineer, I grapple with the challenge of modeling semiconductor fabrication processes using high-dimensional, unlabeled time-series data. Fabrication devices monitor thousands of parameters; there are many nonlinear interactions, such as plasma etching and chemical vapor deposition, where minor fluctuations can propagate into chaotic data patterns. Traditional clustering methods are insufficient for reliably capturing anomalies. The question that drives my work is: how can we embed the underlying physical laws of these processes into AI models, enabling robust anomaly detection without labeled data? To address this, I explored dimension reduction, Koopman analysis, and complex-valued neural networks, developing architectures to identify invariant dynamical system features. As I prepare my first patent, I've recognized the importance of integrating the machine learning theory I've acquired from Caltech with physical meaning—a perspective I hope to expand at Columbia by exploring how such techniques might generalize to broader applications.

Early Research: My early experiences laid the foundation for my interdisciplinary mindset. At the University of Michigan, I conducted statistical analysis of leukemia-inhibitory compounds, mining high-throughput screening data for subtle patterns. As I achieved the goal of finding therapeutic candidates, I gained important insight on the power of mathematical

abstractions in uncovering structure in seemingly noisy data—a lesson that shaped my later work. I presented my findings at the Pathobiology for Investigators, Students, and Academicians Conference and won the Regeneron Scholar Award with my paper.

At USC, I applied unsupervised learning to single-cell sequencing data to study drug resistance mechanisms in cancer cells, drawing broader insights about heterogeneity within biological systems. These early projects helped me recognize AI's potential to generate novel insights by bridging domains.

Thus at Thermo Fisher Scientific, I worked on homography estimation for microscope images, which involved mapping between projective spaces with sparse data; my work was selected for the TFS Innovation Day Award. I worked with small noisy medical datasets and re-implemented several publications, experimenting with mathematical methods, GANs, YOLO, and ultimately a UNet-based solution, but what struck me most was the inherent trade-off between model complexity and interpretability. This realization led me to appreciate the growing importance of physics-informed AI, particularly in contexts like semiconductors and bioinformatics, where predictions must not only be accurate but also explainable. I hope to explore this intersection further at Columbia, leveraging ideas from Dr. David Knowles's work on regulatory genomics and Dr. Itsik Pe'er's research on causal modeling to develop methods that can explain AI's predictions while retaining their power.

Research Interests: In addition to computational biosciences, I've been curious to explore how spatiotemporal embeddings could merge with dimension reduction methods like DMD within ML models. This could allow for efficient representations while retaining critical dynamical features. I am also interested in using Kolmogorov equations for model regularization, as they can embed the dynamics of Markov processes, enabling a greater fidelity of learning. I

would be eager for insights from Dr. Changxi Zheng, CAVE lab, and other faculty on how such techniques could be applied to fields like motion trajectory or acoustics to push the boundaries of explainable AI.

Beyond specific methodologies, I am drawn to the collaborative ethos at Columbia, where researchers tackle complex problems through interdisciplinary partnerships. My experience working across fields has shown me that progress often comes from connecting ideas that initially seem unrelated. For instance, during my time at MIT Lincoln Labs, I applied lessons from bioinformatics to real-time image segmentation, recognizing parallels in how feature hierarchies emerge in both data types. At Bluesky, I combined my expertise in AI and coding to create generative-AI SQL queries and code optimization methods. Such moments of cross-pollination fuel my belief that Columbia's environment—at the nexus of AI and applied sciences—will be instrumental in developing my ideas.

Background and Adversity: While academic rigor has shaped much of my journey, personal challenges have profoundly influenced my perspective. As a student-athlete at Caltech, I excelled academically and athletically, balancing a rigorous 3.9-GPA course load with NCAA basketball. However, my senior year brought profound challenges: Illness hospitalizations marred my transcript while ending my final basketball season. A decade of effort into academics and basketball was abruptly meaningless. Then came the devastating passing of my long-time girlfriend. I was left grappling with immense grief and purposelessness.

At graduation, in the midst of this turmoil, I chose from several opportunities to work on challenging problems in an innovative environment; thus, I started my current research at PerformanceStar. Building remarkable solutions and collaborating with talented scientists and engineers, I was reminded of the joys of community and growth. I understood the resilience required of learning itself—something I had cultivated through all the disparate episodes of my

life, be it research, academics, or athletics. After work, I volunteer as a youth basketball coach and guitar instructor in my hometown. I hope to encourage kids to embrace the joy of learning while giving back to the Silicon Valley tech-centric community that shaped me. From teaching to research, I rediscovered my passion for growth; I felt the urge to learn and stay at the forefront of technology. I believe that a meaningful life is one of continuous learning and contribution—a philosophy I hope to bring to Columbia.

Conc: At Columbia, I envision contributing to groundbreaking AI research in an environment where character growth accompanies personal development. I am particularly excited to engage with the department's diverse expertise, from Dr. Zhou Yu's work in conversational AI to Dr. Hsu's foundational work on transformers. My aim is not only to push the boundaries of existing research but also to pioneer new approaches that connect machine learning with real-world systems, advancing both theory and application. I hope to be part of Columbia's vibrant research community, working with like-minded peers and mentors to develop the AIs of tomorrow and contribute meaningfully to the field. I look forward to joining Columbia, not just to learn, but to share ideas, grow as a researcher, and help shape the future of technology.