

Yihan Shi

(919) 884-0734 | yihan.shiwork@gmail.com | [yihan-shi.github.io](https://github.com/yihan-shi)

Education

Dartmouth College, Thayer School of Engineering

Expected Jun 2026

M.Eng. in Computer Engineering

Duke University

May 2023

B.S., *cum laude*, with High Distinction

Major in Statistics with a Data Science concentration

Minors in Computer Science and Neuroscience

Honors Thesis: *Investigate the impact of Confounding Medical Interventions (CMI) in electronic health records (EHR) data*

Honors and Awards

- Dean's List in Fall 2021, Spring 2022, Fall 2022
- Genentech's Pharmaceutical Technical Development Outstanding Student Award (*awarded to one recipient per university*)
- Undergraduate Travel Award for the 13th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics

Experience

The Brookings Institution, Center on Health Policy

Washington, DC

Senior Research Assistant

Aug 2024 – Present

- Solved a company-wide data bottleneck by independently developing a pipeline to automate downloading, streaming, and structuring messy monthly JSON files, overcoming computing power limitations
- Co-authored and fact-checked research papers and comment letters with economists and lawyers on the Inflation Reduction Act, pharmaceutical import security investigations, and health insurance price transparency
- Managed the full intern hiring process from application review to final decision and provided mentorship and supervision to the selected intern
- Improved data access by web scraping U.S. prescription drug policy information from government websites, SEC reports, and court rulings using Python

Mastercard Data & Services

Arlington, VA

Software Engineer

Aug 2023 – Aug 2024

- Led an innovative research project using variational autoencoders (a type of artificial neural network) to predict counterfactual campaign outcomes in a company-wide AI initiative
- Collaborated with UX designers and product managers to revamp the customer experience, refactoring CSS, HTML, and JavaScript to meet industry-leading accessibility and UX standards
- Developed full-stack web application features to serve enterprise customers such as McDonald's, Walmart, and BofA, allowing large-scale document permission management, lookup, update, upload, and deletion services

Duke University, Department of Statistical Science

Durham, NC

Undergraduate Research Assistant

Aug 2021 – May 2023

- Co-authored internal tutorial on recurrent event analysis to examine instances of alcohol-related complications after bariatric surgeries
- Investigated algorithmic bias of inappropriate proxy variables (e.g., race, gender, and ZIP Code) in Electronic Health Record data using simulation

Global Health Data Science Teaching Assistant

Aug 2022 – Dec 2022

- Held weekly office hours to help more than 70 students understand statistics concepts and answer coding questions
- Guided 2 teams of students in the design of a data analysis project, which involved selecting and cleaning online datasets, generating and testing hypotheses, and training and evaluating statistical models

Carta Healthcare

San Francisco, CA

Data Science Intern

May 2022 – Aug 2022

- Built a client-facing dashboard that tracked 10 key performance metrics and presented it in a live product demonstration

- Created a reusable template that enabled engineers to auto-generate webpages for the pre-training process in a machine learning workflow
- Implemented an automated data-cleaning pipeline that reduced bioinformaticians' housekeeping time by 30%
- Developed infrastructure that ran simulations of patient profiles 100x faster, reducing manual effort by 50%

Duke University School of Medicine

Durham, NC

Summer Fellow

May 2021 – Aug 2021

- Led a team of 3 undergraduate and postdoc students to craft a rare disease genomic data science project end-to-end
- Identified 20+ significant genes out of 35,000 using statistical techniques such as effect size shrinkage and blind dispersion estimation
- Mentored 2 incoming first-year researchers and drafted biomedical data science starter projects

wearehuman.io

Shanghai, China

Digital Marketing Intern

Apr 2020 – Jun 2020

- Designed social media campaigns and analyzed user data to achieve better advertisement rates
- Drafted industry reports on post-COVID social commerce platforms to evaluate the company's potential markets and target audiences
- Organized marketing events such as guest talks with Shanghai University of Finance and Economics

Activities and Leadership

Duke Technology Scholars Program

Durham, NC

Vice President and Scholar

Jan 2021 – May 2023

- Managed 5 technology interest groups, such as Health-tech, Tech for Social Good, and Data Science, with 170+ students
- Recruited 9 small group leaders and mentored them on brainstorming and planning events, creating social media campaigns, and tracking goals effectively
- Set agendas and created outreach worksheets as the organizer for 8 events in 2022, such as "Food for Thought" seminars, technical workshops, and lunch with industry professionals

Statistical Science Majors Union

Durham, NC

Alumni Liaison Officer

Aug 2020 – May 2022

- Organized quarterly career panels and coffee chats to connect current students with Duke alumni in Data Science
- Hosted and streamlined advising sessions, mentorship programs, and networking events for pre-major students

Duke International Association

Durham, NC

Vice President of Policy

Oct 2019 – Aug 2022

- Served as 1 of 3 students on the advisory board of the Duke International Alumni Committee and launched the Duke International Alumni Affinity Group as a platform for global connections
- Collaborated with Duke International Student Center, Counseling & Psychological Services, Visa Services, and Career Center to organize yearly panels on topics such as international students' well-being and Optional Practical Training (OPT)

Publications

- Wosińska, M. E., & **Shi, Y.** (2025). *US drug supply chain exposure to China*. Brookings.
- Frank, R. G., & **Shi, Y.** (2025). *Cumulative net earnings of drugs selected or likely to be selected for negotiation*. Brookings.
- Biswas, D. D., **Shi, Y.**, El Haddad, L. et al. (2024). *Respiratory neuropathology in spinocerebellar ataxia type 7*. JCI Insight, 9(18).
- Roger, A. L., Biswas, D. D., Huston, M. L., ..., **Shi, Y.** et al. (2024). *Respiratory characterization of a humanized Duchenne muscular dystrophy mouse model*. Respiratory Physiology & Neurobiology, 326, 104282.
- Biswas, D. D., Strickland, L. M., Dhindsa, J. S., **Shi, Y.** et al. (2022). *Accumulation of pathological Ataxin-7 in the medulla leads to hypoglossal (XII) Motor Unit Pathology*. The FASEB Journal, 36(S1).

Skills and Interests

- **Languages:** Mandarin Chinese (native), Japanese (intermediate)
- **Technical Skills:** Python, R, SQL, C#, C, Java, JavaScript, CSS, HTML, \LaTeX
- **Interests:** Photography, Cooking, Pilates