Helen Jin

helenjin@seas.upenn.edu • https://helenjin.github.io/
CURRICULUM VITAE - PREPARATION DATE: October 2024

RESEARCH INTERESTS

Keywords: Trustworthy ML/NLP, Explainable AI (XAI), Interpretability, Cognition
I am broadly interested in making AI safer, easier to use, and more helpful to humans. I am also interested in how we can utilize AI to better understand human cognition and intelligence.

EDUCATION

University of Pennsylvania, School of Engineering and Applied Science, Philadelphia, PA   Sep 2020 – Present
Ph.D. Candidate in Computer and Information Science (CIS)
- **Advisor:** Prof. Eric Wong
- Served on the Graduate and Professional Student Assembly (GAPSA) as part of the Executive Board, actively advocating for and working toward improving graduate student life across all twelve schools of Penn
- Perry World House Graduate Associate

Columbia University, Columbia College, New York, NY   Sep 2016 - May 2020
B.A. – Double Major in Mathematics and Computer Science, Concentration in East Asian Studies   GPA*:* 3.72 / 4.00
- **Computer Science Track: Intelligent Systems**
- **Advisor:** Prof. Tian Zheng
- **Relevant Coursework**: Data Structures in Java, Advanced Programming in C/C++, Discrete Mathematics, Fundamentals of Computer Systems, Computer Science Theory, Artificial Intelligence, Machine Learning, Natural Language Processing, Computer Vision, Computation and the Brain, Advanced Spoken Language Processing, Databases, Linear Algebra, Modern Algebra I + II, Modern Analysis I + II, Topology, Calculus-Based Statistics, Probability Theory
- **Honors/Awards**: Dean's List

Girls Who Invest, The Wharton School, University of Pennsylvania, Philadelphia, PA   May 2018 – Aug 2018
Summer Intensive Program Scholar
- Completed highly selective, rigorous 4-week training program focused on core investment concepts and skills taught by leading business school professors and investment professionals; upon completion, earned a paid 6-week internship in asset management
- Coursework and case studies included accounting, valuation, financial modeling, asset allocation and presentation skills

Stuyvesant High School, New York, NY   Sep 2012 – Jun 2016
Advanced Regents Diploma   GPA*:* 4.0 / 4.0
- **Activities**: Girls' Varsity Swimming, Lifeguard, Stuyvesant Spectator, Math Team, CS Dojo and Writing Center Tutor
- **Honors/Awards**: AP National Scholar; National Merit Finalist; ARISTA National Honor Society; National Latin Society; National Junior Classical League; National Latin Exam Awards: Maxima Cum Laude; Award of Distinction in Mathematics

PUBLICATIONS

[7] "Certifiably Robust Evaluation of Feature Attributions via Boolean Influences." **Helen Jin**, Weiqiu You, Eric Wong. *In Submission.*

[6] "FIX: A Benchmark for Features Interpretable to eXperts." **Helen Jin**, Shreya Havaldar, Chaehyeon Kim, Anton Xue, Weiqiu You, Helen Qu, Marco Gatti, Daniel A Hashimoto, Bhuvnesh Jain, Amin Madani, Masao Sako, Lyle Ungar, Eric Wong. [arXiv] [Blog] [Website]

[5] "Linguistic Properties of Truthful Text Generation." Bruce W. Lee, Benedict Florance Arockiaraj, **Helen Jin**. In *TrustNLP: Third Workshop on Trustworthy Natural Language Processing at ACL 2023.* [arXiv]

[4] "Automatically Generated Summaries of Video Lectures Enhance Students' Learning Experience." Hannah Gonzalez*, Jiening Li*, **Helen Jin***, Jiaxuan Ren*, Hongyu Zhang*, Ayotomiwa Akinyele*, Adrian Wang, Eleni Miltsakaki, Ryan Baker, Chris Callison-Burch. In *The 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) at ACL 2023.* [Paper]

[3] "Generic Temporal Reasoning with Differential Analysis and Explanation." Yu Feng, Ben Zhou, Haoyu Wang, **Helen Jin**, Dan Roth. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL).* 2023. [arXiv]

[2] "Large-scale, image-based tree species mapping in a tropical forest using artificial perceptual learning." Tang, C, Uriarte, M, **Jin, H**, Morton, DC, Zheng, T. *Methods Ecol Evol*. 2021; 12: 608– 618. [Paper]

[1] "Artificial Perceptual Learning: Image Categorization with Weak Supervision." Chengliang Tang, Maria Uriarte, **Helen Jin**, Douglas Morton, Tian Zheng. 2021. [arXiv]

*Indicates equal contribution

WORK EXPERIENCE

BrachioLab, University of Pennsylvania, Philadelphia, PA   Sep 2023 – Present
Research Assistant
● PI: Prof. Eric Wong
● Working on problems in Explainable Artificial Intelligence (XAI), specifically the evaluation of explanations

Cognitive Computation Group, University of Pennsylvania, Philadelphia, PA   Sep 2020 – Sep 2023
Research Assistant
● PI: Prof. Dan Roth
● Worked on Natural Language Understanding (NLU) problems, including Complex Event Annotation Project and The State of the Art of Semantic Role Labeling (SRL) Systems

University of New York-Herbert H. Lehman College, Virtual   Jun 2020 – Aug 2020
PTS3 Data Science Mentor-Instructor
● Pathways to Student STEM Success Program (PTS3)
● Supervised the capstone projects of four undergraduate students who previously had little/no prior experience with data science
● Mentored and taught students relevant technology for their projects such as Python, R, Google Colab

Department of Statistics, Columbia University, New York, NY   Jun 2019 – Aug 2020
Undergraduate Researcher
● PI: Prof. Tian Zheng
● Worked on a research project that seeks to develop an artificial perceptual learning framework that can aid semi-supervised learning problems, specifically with a focus on image classification on large images of tree canopies with a small labeled tree species data set
● Used various statistical and deep learning methods primarily in Python, including TensorFlow and Keras packages

LionBase, LLC, New York, NY   Dec 2018 – Sep 2019
Client-Facing Team Member
● LionBase is a new student-led data science and product development group at Columbia University
● Worked in a team of six to solve real-world industry problems related to data analytics, ML, NLP, and statistical analysis

- Spring 2019: Recruitment Platform for Early Stage Data Science Talent – my team collaborated with an executive search firm to design, develop, and implemented from scratch a platform that matches data science talent and company based on various technical and non-technical tests.

Data Science Institute, Columbia University, New York, NY   Mar 2019 – May 2019
Data for Good (DFG) Scholar
- Worked in a team in collaboration with Okimo, a startup using eye tracking technology to help identify and aid individuals' reading skills especially targeting younger age groups in low resource communities in developing countries such as Paraguay

Grantham, Mayo, van Otterloo & Co. LLC (GMO), Boston, MA   Jul 2018 – Aug 2018
Quantitative Equity Analyst Intern
- Using MATLAB and SQL, created a working tool to measure the relationship between institutional investor concentration and downside risk in equities, and modeled risk in global equity markets

Computational and Systems Biology, Memorial Sloan Kettering Cancer Center, New York, NY   May 2017 – Aug 2018
Undergraduate Researcher
- PI: Prof. Dana Pe'er
- Developed computational methods and packages to analyze single cell RNA-seq data using Python and R

TEACHING EXPERIENCE
_____

I have a strong love for teaching and mentoring. I have TA-ed over 10+ courses through the years simply because I find teaching incredibly rewarding and I love to give back the knowledge I have learned. I hope to continue to foster this as I continue in my career, both in formal and informal capacities.

University of Pennsylvania, Philadelphia, PA
- **Teaching Assistant**, Mathematics of Machine Learning [Fall 2024]
- **Teaching Assistant**, Artificial Intelligence [Fall 2021]
  - This is the largest course the university offers. I managed 2 sections of the course simultaneously and single handedly assigned homework pairs.
- **Teaching Assistant**, Girls Who Invest Summer Intensive Program @ The Wharton School [Summer 2023]

University of New York-Herbert H. Lehman College, New York, NY
- **Data Science Mentor-Instructor**, Pathways to Student STEM Success Program (PTS3) [Summer 2020]

Columbia University, New York, NY
- **Teaching Assistant,** Natural Language Processing [Fall 2019, Spring 2020, Summer 2020]
- **Teaching Assistant,** Linear Algebra [Spring 2019]
- **Teaching Assistant,**Calculus IV [Spring 2018, Summer 2019]
- **Teaching Assistant,** Calculus III [Fall 2018, Summer 2019]
- **Teaching Assistant,** Calculus I [Fall 2017]
- **Teaching Assistant,** Calculus-Based Intro to Statistics [Fall 2017]

Other
- **Private Tutor** for Natural Language Processing course at Columbia University [Spring 2021]
- **Private Tutor** for elementary school children, with specific exposure to computer science concepts and coding [Summer 2020]
- **In-Person Private Tutor** for Mathematics and Chemistry subjects, high school level [Feb 2017 – Jun 2017]

- **One-to-One Tutoring Volunteer** - Volunteered weekly to tutor and mentor 6-12 year olds in Harlem community in NYC [2016-2019]

SERVICE ACTIVITIES

- **Reviewer** for: ACL, China National Conference on Computational Linguistics (CCL)
- **Organizer** of the Computational Linguistics (CLunch) seminar at the University of Pennsylvania (Spring 2023)
- **Conference Volunteer** for NAACL
- **CIS Department PhD Mentorship Program Mentor + Event Coordinator Volunteer** (2022-2023, 2023-2024)

University of Pennsylvania, Philadelphia, PA
Representative for:
- **University Council** [Academic Year 2021-2022, Academic Year 2022-2023]
- **University Committee on Honorary Degrees** [Academic Year 2021-2022]
- **Student Advisory Group for the Environment (SAGE)** [Academic Year 2021-2022]
- **University Committee of Academic Affairs** [Academic Year 2022-2023]
- **Ivy+ Summit Conference** [Academic Year 2021-2022 at University of Pennsylvania, Academic Year 2022-2023 at Columbia University]

SELECTED LEADERSHIP EXPERIENCE

Graduate and Professional Student Assembly (GAPSA), University of Pennsylvania, Philadelphia, PA
https://gapsa.upenn.edu

Director of Cultural Programming	May 2024 – Present
Vice President of Operations	May 2022 – May 2023

- Directly oversaw and managed the Operations Division officers
- Solely managed a ~$100K budget
- Served as primary liaison to student governments of each of the graduate and professional schools
- Worked on Pilot GAPSA Meal Swipe Program that passed in the General Assembly– $50,000 of the graduate student government budget was used to provide free meal swipes to students facing food insecurity (resolution)

Director of Logistics	May 2021 – May 2022

Girls Who Code (GWC) at Columbia University, New York, NY   Feb 2017 – Mar 2020
https://outreach.engineering.columbia.edu/content/girls-who-code

President
- Oversaw high school outreach and recruitment for program that holds weekly CS classes for ~50 high school girls each semester
- Planned on-site visits, created fundraisers, and reached out to companies and local businesses for sponsorship
- Worked with other Executive board members to better improve program
- Managed ~30 people on the Managing Board (Finance, Programming, Publicity, HS Recruitment)
- Previously served as **Vice President of External Affairs**, and before that, **Sponsors and Finance Team**

One-to-One Tutoring at Columbia University, New York, NY   Oct 2016 – Sep 2019
https://one2onetutoring.wordpress.com/

Treasury Coordinator, Head Coordinator
- Allocated and managed annual $2,000 budget to various on-campus events from weekly meetings to fundraisers
- Supervised and oversaw volunteers during tutoring sessions each semester to ensure attendance
- Worked with other coordinators and Community Impact staff to discuss issues and implement solutions to improve program

● Volunteered weekly to tutor and mentor 6-12 year old individual throughout academic year

Smart Woman Securities (SWS), Columbia Chapter, New York, NY   Sep 2017 – May 2018
Senior Research Analyst
● Led a team of SWS Analysts researching the Consumer industry
● Conducted background research, analyzed and formulated investment opinions on companies to make real-life investment recommendations

SELECTED OTHER PROJECTS

Semantic Role Labeling (2022)
● Research on predicate and argument identification and classification

Complex Events (2022)
● Research on complex event extraction and knowledge graphs

Persistent Homology (2019)
● Created a visualization tool for the persistent homology in Python for my Math seminar on Elementary Applied Topology

SCAnalysis (2018)
● Built a Python package and Jupyter notebook to analyze single cell RNA-seq data that incorporates MAGIC and Wishbone technologies, with additional features (including Palantir). Wishbone is an algorithm to align single cells from differentiation systems with bifurcating branches. MAGIC (Markov-Affinity Based Graph Imputation of Cells) is an interactive tool to impute missing values in single-cell data and restore the structure of the data.

HONORS & AWARDS
● **Penn Wharton Innovation Fund Validation Award** [Dec 2023]
● **Dean's List**, Columbia University
● **Rewriting the Code (RTC) Fellow** [2017- present]
● **Science Technology Engineering Program (STEP) Award**, Columbia University [2017]

SKILLS & INTERESTS
Technologies: Java, Python, MATLAB, SQL, R, C/C++, Git, Jupyter Notebook, LaTeX , HTML/CSS/JavaScript, Coq

Languages: English, Korean, Mandarin Chinese, minimal Japanese and Spanish

Personal Interests: Psychology, Philosophy, Music, Visual Arts, Swimming, Yoga, Traveling, Hiking