Jonathan P. Chang

jpc362@cornell.edu | cs.cornell.edu/~jpchang | (408) 717-1539

Summary

I am passionate about teaching and mentorship at the undergraduate level. I have taught one undergraduate CS course as co-instructor of record, additionally TA'd six CS courses across nine semesters of my undergraduate and graduate careers, and designed and taught various data science workshops. I have also been a research mentor for six undergraduate students.

My research involves working with real online communities to design and evaluate computational tools for content moderation and community governance. It has been published in top venues for both natural language processing (ACL, EMNLP) and computational social science (WWW, CSCW).

Education

Cornell University 2017–Present

PhD in Computer Science
Advisor: Cristian Danescu-Niculescu-Mizil
Expected Graduation: May 2024

Harvey Mudd College 2013–2017

Bachelor of Computer Science
Dean's List: All Eligible Semesters
Graduated With High Distinction

Teaching Experience

Courses Taught

CS4300 - Information and Language Spring 2023

Cornell University
This was a 200-person course open to both computer science and information science undergraduates, covering selected topics in information retrieval, machine learning, and natural language processing. As co-instructor of record, I updated and taught lectures on the machine learning topics, reorganized some assignments to better spread out the workload, designed exam questions, and oversaw the development of new infrastructure for the course project to make it more approachable for students with less technical background.

Workshops Taught

Introduction to Python Bootcamp Summer 2022

Cornell Center for Social Sciences
I co-developed and co-instructed this brand-new workshop designed to introduce Python to social science researchers with little to no prior programming experience. It covers the basics of the Python language, explains key debugging skills, and introduces core data science packages.

Using APIs for Social Science Research Spring 2022

Cornell Center for Social Sciences
I co-developed this brand-new workshop that explains what APIs are and demos how to use an API to fetch different kinds of social media data.

Introduction to Git and GitHub Fall 2021

Cornell Center for Social Sciences
This workshop covers the basics of version control and how it can both help researchers keep their code organized and promote reproducibility in research. I updated it with new visual explanations of key Git concepts to make it more approachable for first-time Git users.

Courses TA'd

CS4700 - Foundations of Artificial Intelligence, Cornell University Fall 2017

CS151 - Artifical Intelligence, Harvey Mudd College Spring 2017

CS140 - Algorithms, Harvey Mudd College Fall 2017

CS131 - Programming Languages, Harvey Mudd College Spring 2016

CS70 - Data Structures and Program Development, Harvey Mudd College — Fall 2014-16
Head TA during the Fall 2015 semester

CS60 - Principles of Computer Science, Harvey Mudd College — Spring 2014

Publications

Jonathan P. Chang, Charlotte Schluger and Cristian Danescu-Niculescu-Mizil. "Thread With Caution: Proactively Helping Users Assess and Deescalate Tension in Their Online Discussions". Proceedings of CSCW. — 2022

Charlotte Schluger, **Jonathan P. Chang**, Cristian Danescu-Niculescu-Mizil and Karen Levy. "Proactive Moderation of Online Discussions: Existing Practices and the Potential for Algorithmic Support". Proceedings of CSCW. — 2022

Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang and Cristian Danescu-Niculescu-Mizil. "ConvoKit: A Toolkit for the Analysis of Conversations". Proceedings of SIGDIAL (System Demos). — 2020

Jonathan P. Chang, Justin Cheng, and Cristian Danescu-Niculescu-Mizil. "Don't Let Me Be Misunderstood: Comparing Intentions and Perceptions in Online Discussions". Proceedings of The Web Conference (WWW). — 2020

Jonathan P. Chang and Cristian Danescu-Niculescu-Mizil. "Trouble on the Horizon: Forecasting the Derailment of Online Conversations as they Develop". Proceedings of EMNLP. — 2019

Jonathan P. Chang and Cristian Danescu-Niculescu-Mizil. "Trajectories of Blocked Community Members: Redemption, Recidivism and Departure". Proceedings of The Web Conference (WWW). — 2019

Liye Fu, **Jonathan P. Chang** and Cristian Danescu-Niculescu-Mizil. "Inferring Advice-Seeking Intentions From Personal Narratives: A Cloze Test Formulation". Proceedings of NAACL. — 2019

Justine Zhang, **Jonathan P. Chang**, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Nithum Thain, Dario Taraborelli. "Conversations Gone Awry: Detecting Early Signs of Conversational Failure". Proceedings of ACL. — 2018

Jonathan P. Chang and Stefan Scherer. "Learning Representations of Emotional Speech with Deep Convolutional Generative Adversarial Networks". Proceedings of IEEE ICASSP. — 2017

Honors and Awards

"Class of '94" Computer Science Department Award — 2017
Harvey Mudd College
Awarded to top 3 graduating computer science undergraduates on the basis of coursework, research, and service.

Clinic Award — 2017
Harvey Mudd College
Awarded for outstanding senior capstone project.

Departmental Honors in Humanities, Social Sciences, and the Arts — 2017
Harvey Mudd College
Awarded for academic performance in courses in humanities, social sciences, and the arts.

Leadership

Mentorship Programs

Data Science Fellow — 2021-2023
Cornell Center for Social Sciences
In addition to designing and teaching workshops (listed above), I also provided one-on-one consultations to researchers to help them with data science related tasks.

Peer Academic Liaison 2015–2017
Harvey Mudd College
Connected Harvey Mudd students to academic resources and provided support, especially for first-year students.

Departmental Service

CS Department "Czar Czar" 2019–2023
Cornell University
The "czars" system is the CS department's informal program for student volunteer ("czar") positions to organize events such as lunch seminars and social outings. As "czar czar", I was in charge of recruiting and managing the "czars".

Talks and Events

Invited Talks

Trajectories of Blocked Community Members: Redemption, Recidivism & Departure June 2019
Wikimedia Research Showcase

Conversations Gone Awry: Detecting Early Signs of Conversational Failure June 2018
Wikimedia Research Showcase

Invited Events

Breakthroughs in AI Workshop Fall 2019
Microsoft Research

Media Coverage

How AI Tools Could Help Make Online Discussions Healthier 2023
The Academic Minute / Inside Higher Ed

Regret being hostile online? AI tool guides users away from vitriol 2023
Cornell Chronicle

Machine learning could stop an online war of words before it starts 2018
MIT Technology Review

Machine learning is helping computers spot arguments online before they happen 2018
The Verge

Scientists are building a detector for conversations likely to go bad 2018
Fast Company

Predicting when online conversations turn toxic 2018
Cornell Chronicle

Industry Experience

Core Data Science Intern Summer 2019
Facebook
Worked with CDS researcher Justin Cheng on understanding the role of intentions and perceived intentions in public conversations.

Software Engineering Intern Summer 2017
Proofpoint Inc.
Developed a machine learning pipeline for quick filtering of suspicious URLs in emails.

Intern Software Analyst and Developer Summer 2015
Intentional Software Corporation
Leveraged Intentional's application platform to create a unique color editor tool allowing users to manipulate 3d representations of color space.

Software Engineering Intern Summer 2014

Smith Micro Software

Implemented a content browser for Poser Pro (Smith Micro's 3d animation tool) using JavaScript. Added texture processing and multi-animation support to Poser Pro's FBX file format exporter using C++. Created a demo game using the Unity engine to demonstrate Poser Pro's Unity support.

Software and Datasets

ConvoKit

An open-source Python NLP package implementing multiple conversational analysis tools. I led a redesign effort in 2020 to improve user friendliness and flexibility for the 2.0 release, and again in 2021-22 to add support for a new optional database storage backend. In addition to leading these efforts, I also continue to be an active contributor of code, documentation, and ideas.

https://convokit.cornell.edu

https://github.com/CornellNLP/ConvoKit

CRAFT

A beginner-friendly reimplementation of the CRAFT language model from my EMNLP 2019 paper "Trouble on the Horizon", written in and compatible with the popular PyTorch neural network framework.

https://github.com/jpwchang/CRAFT

Conversations Gone Awry Datasets

Datasets of conversations from Wikipedia Talk Pages and Reddit's "ChangeMyView" debate forum, annotated for the presence of personal attacks and hostility. Over 10,000 annotated conversations in total, distributed as part of ConvoKit.

https://convokit.cornell.edu/documentation/awry.html

https://convokit.cornell.edu/documentation/awry_cmv.html

Skills

Python	C/C++	OCaml
PyTorch	C#	Perl
SpaCy	Java	Prolog
NLTK	JavaScript	Racket
scikit-learn	PHP	Ruby
Numpy / Scipy	Haskell	SQL
Pandas	HTML / CSS	

Relevant Coursework

Artificial Intelligence	Algorithms	Autonomous Vehicles
NLP	Computability and Logic	NLP & Social Interaction
Neural Networks	Probability and Statistics	Information Networks
Computer Vision	Image Processing	