

# Will Epperson

I'm a Ph.D. student in the HCII at CMU advised by Dominik Moritz and Adam Perer.

My research interests lie in **developing interactive data science tools for both experts and non-experts**. I am interested in how we can help analysts understand their data through interactive visualization, recommended analysis, and models. My commonly used research methods include building stand-alone systems for data analysis, building extensions to existing tools like Jupyter, and running human studies experiments to evaluate these systems.

🏠 [willepperson.com](http://willepperson.com)

✉ [willepp@cmu.edu](mailto:willepp@cmu.edu)

📄 [CV PDF](#)

🐦 [@w\\_epperson](https://twitter.com/w_epperson)

🐙 [@willeppy](https://github.com/willeppy)

👤 [Google Scholar](#)

🌐 [LinkedIn](#)

## Education

August 2020 - Present

### Ph.D. in Human Computer Interaction

Carnegie Mellon University

*Advisors:* Dominik Moritz, Adam Perer

*Sample Coursework:* HCI Process and Theory, Computational Medicine, Human Judgement and Decision Making

August 2016 — May 2020

### B.S. in Computer Science

Georgia Institute of Technology

GPA: 4.0, Summa Cum Laude, threads in Intelligence and Modeling/Simulation

*Sample Coursework:* Machine Learning, Deep Learning, Computer Vision, Computer Architecture, Algorithms, Computer Simulation, Information Visualization

## Publications

### FairVis: Visual Analytics for Discovering Intersectional Bias in Machine Learning

Angel Cabrera, Will Epperson, Fred Hohman, Minsuk Kahng, Jamie Morgenstern, Duen Horng (Polo) Chau

FairVis is a Visual Analytics system that allows users to audit their machine learning models for intersectional bias by exploring the model's performance on various user-specified and recommended subgroups in a dataset.

*IEEE Conference on Visual Analytics Science and Technology (VAST). Vancouver, Canada, 2019.*

🔗 [Project](#) ▶ [Demo](#) 📄 [PDF](#) 📖 [Blog](#) 🎥 [Recording](#) 📽 [Slides](#) 💻 [Code](#) 📄 [BibTeX](#)

### RECAST: Interactive Auditing of Automatic Toxicity Detection Models

Austin P. Wright, Omar Shaikh, Haekyu Park, Will Epperson, Muhammed Ahmed, Stephane Pinel, Diyi Yang, Duen Horng (Polo) Chau

RECAST is an interactive tool that allows users to audit toxicity detection models with their own input text and suggests alternative wordings for detected toxic speech.

*The eighth International Workshop of Chinese CHI. 2020.*

🔗 [Project](#) 📄 [PDF](#) 📄 [BibTeX](#)

## Talks

### FairVis

May 2019

IEEE Visualization Conference

# Honors and Awards

	2019
PURA: President's Undergraduate Research Award \$1500 research grant to continue work on FairVis project	
	2016
Stamps President's Scholarship Full ride scholarship given to 40 incoming freshman at Georgia Tech	

# Research Experience

	August 2020 - Present
<b>Carnegie Mellon University</b> , Pittsburgh, PA <i>Graduate Researcher, Data Interaction Group (DIG)</i> Advisor: Dominik Moritz, Adam Perer Member of the DIG research group, working on novel data visualizations, ML interpretation techniques, and interactive data systems. <i>Relevant Skills:</i> Python, Javascript	
	January 2019 - May 2020
<b>Georgia Institute of Technology</b> , Atlanta, GA <i>Undergraduate Researcher, Polo Club of Data Science</i> Advisor: Duen Horng (Polo) Chau Member of the Polo Club of Data Science working on novel data visualizations to find fairness issues in Machine Learning models <i>Relevant Skills:</i> Python, Javascript	
	January 2018 - May 2019
<b>Georgia Institute of Technology</b> , Atlanta, GA <i>Undergraduate Researcher, Automated Algorithm Design</i> Advisor: Jason Zutty, Greg Rohling Worked on EMAD algorithm design engine to implement sentiment analysis pipeline to analyze news articles to aid in predicting stock price movements using genetic algorithms. Led project to visualize the genetic algorithm evolution process. <i>Relevant Skills:</i> Python, Javascript	

# Industry Experience

	Summer 2021
<b>Microsoft Research</b> , Redmond, WA <i>Research Intern, VIDA Group</i> Mentor: Steve Drucker, Rob DeLine Research intern working on data science tools. Lead project around reuse and sharing in data science, in submission to ICSE 2022. Also involved with project around visualizing data frame differences in submission to CHI 2022. <i>Relevant Skills:</i> Python, Typescript	
	Summer 2019
<b>Point72 Asset Management</b> , New York, NY <i>Data Analytics Intern, Market Intelligence Group</i> Mentor: Trevor Rempel Worked as Data Scientist in alternative data space to clean, model, and understand large datasets <i>Relevant Skills:</i> Python, Distributed Computing in Spark	
	Summer 2018
<b>Ultimate Software</b> , Weston, FL <i>Software Development Intern, Innovation Strategies Team</i> Mentor: Joseph Cutrono Designed and developed Slack app to integrate with the UltiPro HR management tool. App published to Slack app store. <i>Relevant Skills:</i> Typescript, REST API development	
	Summer 2015

**The Home Depot**, Atlanta, GA

*Software Development Intern*

Developed web app for tracking candidate progress throughout hiring process for internal HR use.

*Relevant Skills:* Java, HTML/CSS/Javascript

## Mentees

During my PhD, I have had the pleasure of mentoring the following undergraduate students on research projects.

Summer 2021+

**Leijie Wang**

Visualization recommendation for python in notebooks using history

Fall 2021+

**Asad Sheikh**

Visualization recommendation for SQL using history

## Teaching

August 2017 - December 2018

**Undergraduate Teaching Assistant**

*Georgia Institute of Technology, Atlanta, GA*

Intro to Database Systems (CS 4400), Instructor: Monica Sweat

Designed projects, held office hours and graded for relational databases class.

## Leadership & Activities

January 2019 - May 2020

**Student Ambassador**

*Georgia Institute of Technology Alumni Association*

Serve as official representative of the Institute at events/tours for alumni, prospective students, and special guests.

January 2018 - May 2019

**Executive Board Member -- Threads Co-chair**

*Stamps Scholars National Convention 2019*

Executive board member of Stamps Scholars National Convention, a 3-day conference with over 700 student attendees. Responsible for 20-person committee that plans and coordinates the different content threads of the convention.

## Sample Projects

December 2018

**ICLR'19 Reproducibility Challenge**

Implemented the architecture and reproduced results for an ICLR'19 submission using GANs to de-correlate sensitive data, titled Generative Adversarial Models for Learning Private and Fair Representations.

January 2018 - December 2018

**Atlanta Crime Map**

Led a team as part of Data Science Club at GT to analyze and visualize crime data on and around GT's campus to provide insight into crime frequency and details for GTPD.

February 2018

**Citadel Datathon**

Created supervised learning system to predict dangerous road areas to help make cities safer at Citadel and Correlation One's Datathon at Georgia Tech. Team placed top 3.

February 2017 - December 2017

**RECONSO Research**

Developed interface between core state machine and battery systems on Avionics team of student-led cube satellite project. Learned C while working on project.

# Skills

**Programing Languages:** Python (Advanced), Javascript/Typescript (Intermediate), HTML (Intermediate), SQL (Intermediate), Java (Intermediate), C (Basic)

**Toolkits, Frameworks, Software:** Pytorch, Scikit-learn, Git, VegaLite, D3, Tableau, MacOS, Windows, Linux

**Natural Languages:** English (Native), Spanish (Advanced)