

Shan Jiang | Résumé

660-674 (#86), Interdisciplinary Science and Engineering Complex, 805 Columbus Ave, Boston, MA 02120

☎ (+1) 781-502-8799 | ✉ sjiang@ccs.neu.edu | 🏠 shanjiang.me | 🖨️ printfoo | 🌐 [shan-jiang](https://www.linkedin.com/in/shan-jiang)

Education

Northeastern University

Ph.D. in Computer Science

| Advisor: Christo Wilson

Boston, MA

Sep 2016 - 2021 (Expected)

Beijing University of Posts and Telecommunications

B.B.A. in Management Information Systems

| Rank: 1/46 GPA: 92.5/100

Beijing, China

Sep 2012 - Jul 2016

Experience

Google

Software Engineer Intern @ Fact Check Team, Google AI

New York, NY

Jun 2019 - Aug 2019

- Project: ClaimReview markup (e.g., claim, claimant and verdict) extraction from fact-check articles.
- Explored several task formulation possibilities, e.g., language generation under encoder-decoder frameworks, and eventually formulated the task as a sequence tagging problem and conducted several experiments by modifying and fine-tuning BERT models.
- **Productionized** 5K+ lines of codebase with test files and technical documentation.
- Prepared a paper with additional data exploration and model comparison to be submitted to [WWW'20](#).

Dataminr

Research Intern @ AI and Data Science Team

New York, NY

Feb 2019 - Apr 2019

- Project: Crisis sub-event (e.g., burning road after a wildfire) detection on social media for emergency management.
- Built a pipeline model that first scans the Twitter firehose and parses Tweets to dependency trees, then traverses to extract connected noun-verb pairs (e.g., home-burn, house-destroy), and finally clusters similar pairs as sub-events.
- Case-studied California wildfires to understand the temporal cascading (e.g., fire→smoke→pollution) of sub-event networks.
- Published a paper at the AI for social good workshop, i.e., [AISG'19@ICML](#).

Northeastern University

Research Assistant @ Khoury College of Computer Sciences

Boston, MA

Sep 2016 - Present

- Research areas: computational journalism, computational social science, algorithm auditing, information quality.
- Collected and analyzed TB-sized social media and search engine data under the Spark/Hadoop framework.
- Applied statistical (e.g., regression) and causal (e.g., propensity score matching) models for hypothesis testing.
- Leveraged empirical observations to build natural language processing and machine learning pipelines to identify misinformation and linguistic bias in human-generated content (e.g., news, comments), particularly under algorithmic curation (e.g., ranking, personalization).
- Published **award-winning** papers at top web and HCI conferences, e.g., [WWW'18-19](#), [ICWSM'19](#), [FAT*19](#), [CSCW'18](#).

National University of Singapore

Research Assistant @ School of Computing

Singapore

Dec 2015 - May 2016

- Project: Economic modeling of Bitcoin mining under risk aversion assumptions.

Beijing University of Posts and Telecommunications

Research Assistant @ State Key Lab of Networking and Switching Technology

Beijing, China

Oct 2013 - Dec 2015

- Project: Game-theoretic modeling of overlay networks and traffic engineering.
- Published papers at computer network and system conferences, e.g., [GlobeCom'15](#), [LCN'15](#), [ICPADS'14](#).

Selected Publications

Bias Misperceived: The Role of Partisanship and Misinformation in YouTube Comment Moderation

ICWSM'19

[Shan Jiang](#), Ronald E Robertson and Christo Wilson

outstanding analysis paper: 0.4% | acceptance rate: 21%

Crisis Sub-Events on Social Media: A Case Study of Wildfires

AISG'19@ICML

[Shan Jiang](#), William Groves, Sam Anzaroot and Alejandro Jaimes

oral presentation: 18%

Who's the Guinea Pig? Investigating Online A/B/n Tests in-the-Wild

FAT*19

[Shan Jiang](#), John Martin and Christo Wilson

acceptance rate: 24%

Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media

CSCW'18

[Shan Jiang](#) and Christo Wilson

acceptance rate: 26%

Auditing Partisan Audience Bias within Google Search

CSCW'18

Ronald E Robertson, [Shan Jiang](#), Kenneth Joseph, Lisa Friedland, David Lazer and Christo Wilson

honorable mention: 2.7% | acceptance rate: 26%

On Ridesharing Competition and Accessibility: Evidence from Uber, Lyft, and Taxi

WWW'18

[Shan Jiang](#), Le Chen, Alan Mislove and Christo Wilson

acceptance rate: 15%

Skills

Languages

Python, Java, Scala, C/C++, R, SQL, HTML/CSS, JavaScript, Bash

Tools

Spark, Hadoop, TensorFlow, PyTorch, Linux, MatLab, Git

September 23, 2019

Shan Jiang