

# Shan Jiang | Curriculum Vitae

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

☎ (+1) 781-502-8799 | ✉ [sjiang@ccs.neu.edu](mailto:sjiang@ccs.neu.edu) | 🏠 [shanjiang.me](http://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

### Facebook

Ph.D. ML Intern @ Illicit Trade Team, Dangerous Content

Seattle, WA

Jun 2020 - Sep 2020

- Project: Weak supervision and multitask learning for detecting illicit trade content.
- Implemented weak supervision on multimodal models for illicit trade (e.g., prostitution, firearms, drugs) detection by utilizing existing labels and pre-trained models of auxiliary tasks (e.g., user reporting, nudity classifier) for multitask learning.
- Proposed and tested generalizable hypotheses on selecting and ordering auxiliary tasks for weakly supervising models in other problem areas, and shared experiments and results across Facebook's machine learning teams.
- Improved the performance of production models by up to **2%** and **productionized** 5K+ lines of code.

### Google

Ph.D. SWE 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 code with test files and technical documentation.
- Published a paper with additional data exploration and model comparison at **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 and Teaching 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 (e.g., **WWW'18-19**, **ICWSM'19-20**), HCI (e.g., **CSCW'18**) and AI (e.g., **AAAI'20**, **FAT\*'19**) conferences.

### 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**.

## Skills

### Programming Languages

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

### Tools & Platforms

Spark, Hadoop/HDFS, TensorFlow/Keras, PyTorch, Git, Linux

### Deep Learning & NLP

Transformers/BERT, Seq2Seq, RNN/LSTM, Attention, Rationalization, Interpretability, Explainability

### Statistics

Hypothesis Testing, Regression Analysis, Causal Inference

## Publications

<b>Building and Auditing Fair Algorithms: A Case Study in Candidate Screening</b> Christo Wilson, Avijit Ghosh, <a href="#">Shan Jiang</a> , Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel and Frida Polli	<b>FACCT'21</b> acceptance rate: 25%
<b>From Dark to Light: The Many Shades of Sharing Misinformation Online</b> Miriam Metzger, Andrew Flanagan, Paul Mena, <a href="#">Shan Jiang</a> and Christo Wilson	<b>MaC'21</b> journal paper
<b>Modeling and Measuring Expressed (Dis)belief in (Mis)information</b> <a href="#">Shan Jiang</a> , Miriam Metzger, Andrew Flanagan and Christo Wilson	<b>ICWSM'20</b> acceptance rate: 17%
<b>Factoring Fact-Checks: Structured Information Extraction from Fact-Checking Articles</b> <a href="#">Shan Jiang</a> , Simon Baumgartner, Abe Ittycheriah and Cong Yu	<b>WWW'20</b> acceptance rate: 19%
<b>Reasoning about Political Bias in Content Moderation</b> <a href="#">Shan Jiang</a> , Ronald E Robertson and Christo Wilson	<b>AAAI'20</b> invited paper: 100%
<b>Bias Misperceived: The Role of Partisanship and Misinformation in YouTube Comment Moderation</b> <a href="#">Shan Jiang</a> , Ronald E Robertson and Christo Wilson	<b>ICWSM'19</b> <i>outstanding analysis paper: 0.4%</i>   acceptance rate: 21%
<b>Crisis Sub-Events on Social Media: A Case Study of Wildfires</b> <a href="#">Shan Jiang</a> , William Groves, Sam Anzaroot and Alejandro Jaimes	<b>AISG'19@ICML</b> oral presentation: 18%
<b>Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation</b> Ronald E Robertson, <a href="#">Shan Jiang</a> , David Lazer and Christo Wilson	<b>WebSci'19</b> acceptance rate: 24%
<b>Auditing the Partisanship of Google Search Snippets</b> Desheng Hu, <a href="#">Shan Jiang</a> , Ronald E Robertson and Christo Wilson	<b>WWW'19</b> acceptance rate: 18%
<b>Who's the Guinea Pig? Investigating Online A/B/n Tests in-the-Wild</b> <a href="#">Shan Jiang</a> , John Martin and Christo Wilson	<b>FACCT'19</b> acceptance rate: 24%
<b>Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media</b> <a href="#">Shan Jiang</a> and Christo Wilson	<b>CSCW'18</b> acceptance rate: 26%
<b>Auditing Partisan Audience Bias within Google Search</b> Ronald E Robertson, <a href="#">Shan Jiang</a> , Kenneth Joseph, Lisa Friedland, David Lazer and Christo Wilson	<b>CSCW'18</b> <i>honorable mention: 2.7%</i>   acceptance rate: 26%
<b>On Ridesharing Competition and Accessibility: Evidence from Uber, Lyft, and Taxi</b> <a href="#">Shan Jiang</a> , Le Chen, Alan Mislove and Christo Wilson	<b>WWW'18</b> acceptance rate: 15%
<b>Conflicts in Overlay Environments: Inefficient Equilibrium and Incentive Mechanism</b> Jianxin Liao, Jun Gong, <a href="#">Shan Jiang</a> , Tonghong Li and Jingyu Wang	<b>KSII-TIIS'16</b> journal paper
<b>Interactions of Overlays and Traffic Engineering: Equilibrium and Cooperation without Payment</b> <a href="#">Shan Jiang</a> , Jun Gong, Jingyu Wang, Jianxin Liao and Tonghong Li	<b>GlobeCom'15</b> acceptance rate: 35%
<b>Competitive Equilibrium and Stable Coalition in Overlay Environments</b> <a href="#">Shan Jiang</a> , Jianxin Liao, Jun Gong, Jingyu Wang and Tonghong Li	<b>LCN'15</b> acceptance rate: 30%
<b>Combination Feature for Image Retrieval in the Distributed Datacenter</b> Di Yang, Jianxin Liao, Qi Qi, Jingyu Wang, Haifeng Sun and <a href="#">Shan Jiang</a>	<b>ICPADS'14</b> acceptance rate: 30%

## Honors and Awards

<b>Outstanding Analysis Paper</b>	for the top analysis paper at <b>ICWSM'19</b> (1/238)	2019
<b>Honorable Mention</b>	for top papers at <b>CSCW'18</b> (30/1,106)	2018
<b>Dean's Fellowship</b>	for 1st-year Ph.D. students at Northeastern University	2016
<b>National/1st-Class Scholarship</b>	for top 1% students at Beijing University of Posts and Telecommunications	2013 - 2015

## Service

<b>Program Committee</b>	CSCW, ICWSM, AAAI	2021
	ICWSM, WebSci	2020
	ASONAM (Multidisciplinary Track)	2019
<b>Reviewer</b>	CHI, CSCW, ICWSM, AAAI	2021
	CHI, CSCW, ICWSM, WebSci	2020
	CHI, CSCW, ICWSM	2019
	CSCW, WWW	2018