

# 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.
- Integrated 5K+ lines of code [in production](#) and improved precision@recall=90% of production models by up to **10%**.

### 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.
- Integrated 5K+ lines of code [in production](#) with test files and technical documentation and published a paper at [WWW'20](#).

### Dataminr

Research Intern @ AI and Data Science Team

New York, NY

Feb 2019 - Apr 2019

- Project: Crisis sub-event (e.g., road burning after a wildfire) identification on social media for emergency management.
- Built a pipeline model that first scans Twitter and parses each Tweet to a dependency tree, then traverses it to extract connected noun-verb pairs (e.g., home-burn, house-destroy), and finally clusters similar pairs as sub-events.
- Case-studied recorded 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 social science, algorithm auditing, misinformation, fact-checking, etc.
- Collected and analyzed TB-sized social media and news data with the Spark/Hadoop framework.
- Applied statistical (e.g., regression) and causal (e.g., propensity scoring) models for hypothesis testings.
- 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](#) research papers at top conferences for natural language processing (e.g., [ACL'21](#)), computational social science (e.g., [WWW'18/19](#), [ICWSM'19/20](#)), HCI (e.g., [CSCW'18](#)) and AI (e.g., [AAAI'20](#), [FAccT'19/21](#)).

### National University of Singapore

Research Assistant @ School of Computing

Singapore

Dec 2015 - May 2016

- Project: Economic modeling of mining cryptocurrency (e.g., Bitcoin) behaviors under assumptions of risk aversion.

### Beijing University of Posts and Telecommunications

Research Assistant @ Institute of Network Technology

Beijing, China

Oct 2013 - Dec 2015

- Project: Game-theoretic modeling of the competition and cooperation between overlay networks and traffic engineering.
- Published research papers at conferences for networks (e.g., [GlobeCom'15](#), [LCN'15](#)) and systems (e.g., [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, Linear Models, Causal Inference

## Publications

### Structurizing Misinformation Stories via Rationalizing Fact-Checks

Shan Jiang and Christo Wilson

ACL'21

acceptance rate: 21%

### Building and Auditing Fair Algorithms: A Case Study in Candidate Screening

Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel and Frida Polli

FAccT'21

acceptance rate: 25%

### From Dark to Light: The Many Shades of Sharing Misinformation Online

Miriam Metzger, Andrew Flanagin, Paul Mena, Shan Jiang and Christo Wilson

Mac'21

journal paper

### Modeling and Measuring Expressed (Dis)belief in (Mis)information

Shan Jiang, Miriam Metzger, Andrew Flanagin and Christo Wilson

ICWSM'20

acceptance rate: 17%

### Factoring Fact-Checks: Structured Information Extraction from Fact-Checking Articles

Shan Jiang, Simon Baumgartner, Abe Ittycheriah and Cong Yu

WWW'20

acceptance rate: 19%

### Reasoning about Political Bias in Content Moderation

Shan Jiang, Ronald E Robertson and Christo Wilson

AAAI'20

invited paper: 100%

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

Shan Jiang, Ronald E Robertson and Christo Wilson

ICWSM'19

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

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

Shan Jiang, William Groves, Sam Anzaroot and Alejandro Jaimes

AISG'19@ICML

oral presentation: 18%

### Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation

Ronald E Robertson, Shan Jiang, David Lazer and Christo Wilson

WebSci'19

acceptance rate: 24%

### Auditing the Partisanship of Google Search Snippets

Desheng Hu, Shan Jiang, Ronald E Robertson and Christo Wilson

WWW'19

acceptance rate: 18%

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

Shan Jiang, John Martin and Christo Wilson

FAccT'19

acceptance rate: 24%

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

Shan Jiang and Christo Wilson

CSCW'18

acceptance rate: 26%

### Auditing Partisan Audience Bias within Google Search

Ronald E Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer and Christo Wilson

CSCW'18

honorable mention: 2.7% | acceptance rate: 26%

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

Shan Jiang, Le Chen, Alan Mislove and Christo Wilson

WWW'18

acceptance rate: 15%

### Conflicts in Overlay Environments: Inefficient Equilibrium and Incentive Mechanism

Jianxin Liao, Jun Gong, Shan Jiang, Tonghong Li and Jingyu Wang

KSII-TIIS'16

journal paper

### Interactions of Overlays and Traffic Engineering: Equilibrium and Cooperation without Payment

Shan Jiang, Jun Gong, Jingyu Wang, Jianxin Liao and Tonghong Li

GlobeCom'15

acceptance rate: 35%

### Competitive Equilibrium and Stable Coalition in Overlay Environments

Shan Jiang, Jianxin Liao, Jun Gong, Jingyu Wang and Tonghong Li

LCN'15

acceptance rate: 30%

### Combination Feature for Image Retrieval in the Distributed Datacenter

Di Yang, Jianxin Liao, Qi Qi, Jingyu Wang, Haifeng Sun and Shan Jiang

ICPADS'14

acceptance rate: 30%

## Honors and Awards

### Outstanding Analysis Paper

for the top analysis paper at ICWSM'19 (1/238)

2019

### Honorable Mention

for top papers at CSCW'18 (30/1,106)

2018

### Dean's Fellowship

for 1st-year Ph.D. students at Northeastern University

2016

### National/1st-Class Scholarship

for top 1% students at Beijing University of Posts and Telecommunications

2013 - 2015

## Service

### Program Committee & Reviewer

AAAI, CSCW, ICWSM, NeurIPS, WebSci

2021

CHI, CSCW, ICWSM, WebSci

2020

ASONAM, CHI, CSCW, ICWSM

2019

CSCW, WWW

2018