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 | shanjiang.me | printfoo | shan-jiang

Education

Northeastern University — Boston, MA
Ph.D. in Computer Science — *Sep 2016 - 2021 (Expected)*
- Advisor: Christo Wilson

Beijing University of Posts and Telecommunications — Beijing, China
B.B.A. in Management Information Systems — *Sep 2012 - Jul 2016*
- Rank: 1/46 GPA: 92.5/100

Experience

Facebook — Seattle, WA
Ph.D. ML Intern @ Illicit Trade Team, Dangerous Content — *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 — New York, NY
Ph.D. SWE Intern @ Fact Check Team, Google AI — *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 — New York, NY
Research Intern @ AI and Data Science Team — *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 — Boston, MA
Research and Teaching Assistant @ Khoury College of Computer Sciences — *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 — Singapore
Research Assistant @ School of Computing — *Dec 2015 - May 2016*
- Project: Economic modeling of Bitcoin mining under risk aversion assumptions.

Beijing University of Posts and Telecommunications — Beijing, China
Research Assistant @ State Key Lab of Networking and Switching Technology — *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

Modeling and Measuring Expressed (Dis)belief in (Mis)information — ICWSM'20
Shan Jiang, Miriam Metzger, Andrew Flanagin and Christo Wilson — *acceptance rate: 17%*

Factoring Fact-Checks: Structured Information Extraction from Fact-Checking Articles — WWW'20
Shan Jiang, Simon Baumgartner, Abe Ittycheriah and Cong Yu — *acceptance rate: 19%*

Reasoning about Political Bias in Content Moderation — AAAI'20
Shan Jiang, Ronald E Robertson and Christo Wilson — *invited paper: 100%*

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%*

Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation — WebSci'19
Ronald E Robertson, Shan Jiang, David Lazer and Christo Wilson — *acceptance rate: 24%*

Auditing the Partisanship of Google Search Snippets — WWW'19
Desheng Hu, Shan Jiang, Ronald E Robertson and Christo Wilson — *acceptance rate: 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%*

Conflicts in Overlay Environments: Inefficient Equilibrium and Incentive Mechanism — KSII-TIIS'16
Jianxin Liao, Jun Gong, Shan Jiang, Tonghong Li and Jingyu Wang — *impact factor: 0.61*

Interactions of Overlays and Traffic Engineering: Equilibrium and Cooperation without Payment — GlobeCom'15
Shan Jiang, Jun Gong, Jingyu Wang, Jianxin Liao and Tonghong Li — *acceptance rate: 35%*

Competitive Equilibrium and Stable Coalition in Overlay Environments — LCN'15
Shan Jiang, Jianxin Liao, Jun Gong, Jingyu Wang and Tonghong Li — *acceptance rate: 30%*

Combination Feature for Image Retrieval in the Distributed Datacenter — ICPADS'14
Di Yang, Jianxin Liao, Qi Qi, Jingyu Wang, Haifeng Sun and Shan Jiang — *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*
Outstanding Undergraduate	for top undergraduate students in the city of Beijing	*2016*
National Scholarship	for top 1% students at Beijing University of Posts and Telecommunications	*2014*
First-Class Scholarship $\times$**2**	for top 2% students at Beijing University of Posts and Telecommunications	*2013, 2015*

Service

Program Committee	ICWSM, AAAI	*2021*
	ICWSM, WebSci	*2020*
	ASONAM (Multidisciplinary Track)	*2019*
Reviewer	CHI, CSCW, ICWSM, AAAI	*2021*
	CHI, CSCW, ICWSM, WebSci	*2020*
	CHI, CSCW, ICWSM	*2019*
	CSCW, WWW	*2018*