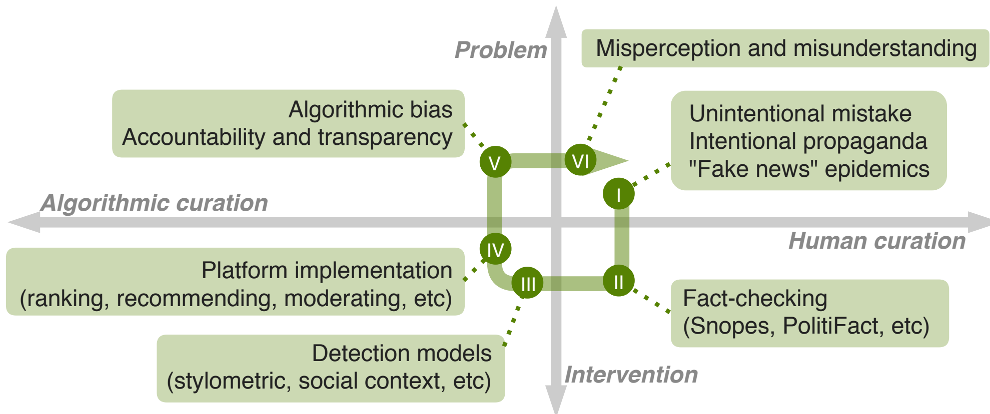
Research Summary | Shan Jiang

660-674, Interdisciplinary Science & Engineering Complex, 805 Columbus Ave, Boston, MA 02120

(+1) 781-502-8799 | sjiang@ccs.neu.edu | shanjiang.me | printfoo | shan-jiang

Human and Algorithmic Curation of Misinformation



The spread of misinformation is a problem as old as communication, but has been recently exacerbated due to the emergence of computer-mediation. To help shed light on this, I mapped the evolution of misinformation onto the human and algorithmic axis (see figure), enabling researchers to pinpoint solutions for it. Using this model, I've examined how misinformation and fact-checking affect social media users' conversational threads, audited algorithmic bias in Google Search, and investigated accusations of biased content moderation against YouTube.

I II III **Human Curated Misinformation and Fact-Checking.** Through computer-mediated communication, e.g., social media platforms, misinformation and fact-checking can impact online users in many ways. In CSCW'18a, I studied how linguistic signals in social media users' conversational threads vary under misinformation and fact-checking. I collected 20K+ fact-check articles from Snopes and PolitiFact, and 2M+ user comments from Facebook, Twitter, and YouTube. Using sentiment analysis, I found that misinformation significantly discourages reasoned conversation by promoting hate speech and aggressive emotional cues in users' discussions. While fact-checking sometimes has positive effects, it also "backfires" on audiences who view it as biased and unreliable. Leveraging these linguistic signals, I built detection models to help the design of more effective social computing systems for misinformation interventions.

IV V **Misinformation and Bias in Algorithmic Intervention.** Algorithmic curated platforms that rank, recommend and moderate information sometimes generate bias and spread misinformation themselves. In CSCW'18b, my collaborators and I looked at potential partisan bias within Google Search. We recruited 187 participants to complete a survey and install a browser extension that enabled us to collect search result from their computers. Leveraging a panel of registered voters, we quantified partisan bias of a website based on how it was shared by users with different party affiliations on social media. Our results showed that the partisan bias of search results depends largely on the queries being conducted, rather than self-reported political ideologies of participants, providing no evidence for "filter bubbles" hypothesis. Our on-going work further investigates this issue for auto-completion (submitted to WWW'19a) and snippets (submitted to WWW'19) generated by search engines.

V VI **Misperception and Misunderstanding of Algorithmic Bias.** Although some concerns about algorithmic bias are grounded, many claims surrounding this issue circle back to human-side misperception and misunderstanding. In my recent work (preparing to submit to ICWSM'19), I investigated one specific claim that "social media is totally discriminating against Republican/Conservative voices (Donald Trump on Twitter)" using YouTube's comment moderation as a lens. I found that although comments under right-leaning videos are moderated more heavily, these comments also correlate to a variety of linguistic signals (e.g., hate speech). After controlling for confounders using a causal propensity score matching model, I found no evidence to support the claim of biased moderation against YouTube. Instead, I found that comments are more likely to be moderated if the video publisher is extremely partisan in either direction (left or right), if the video content is false, and if the comments were posted after fact-checking.

References

Please refer to my **Curriculum Vitae**.