Identifying, Measuring and Contesting Algorithmically Curated Misinformation

Prerna Juneja
University of Washington
Seattle, WA, USA
larst@affiliation.org

ABSTRACT

This research examines the role of algorithms driving the online platforms in surfacing misinformation. Specifically, my dissertation work explores how can we ethically develop scalable audit pipelines to identify, measure and contest algorithmically curated misinformation. I first design experiments to audit and measure online platforms for misinformation across user features, user actions and high impact events. Next, I propose a workflow that combines human and AI capabilities to scale misinformation annotations using a value sensitive design approach. Lastly, I propose to explore how users would like to contest problematic algorithmic outputs and how can online platforms design for algorithmic contestability in scenarios where algorithms expose users to problematic content.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Content ranking**; **Web search engines**; **Web crawling**; • **Human-centered computing** → **Empirical studies in HCI**; *Empirical studies in collaborative and social computing*.

KEYWORDS

Algorithmic bias, misinformation audit, fact-checking, algorithmic explainability, algorithmic contestability

ACM Reference Format:
Prerna Juneja. 2021. Identifying, Measuring and Contesting Algorithmically Curated Misinformation. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion), October 23–27, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3462204.3481788

1 INTRODUCTION

Search engines are the primary gateways of information. However, they are not designed to take into account the credibility of the information while presenting it to us. Dependence on the search systems in addition to our deep rooted trust in their results have made us susceptible to their impact in critical ways. For example, several people ended up believing that the Earth is flat after watching recommended videos on Youtube [8]. If citizens fail to view inaccurate results with a critical eye, or if the search interfaces themselves incentivize the public towards more conspiratorial content (for e.g., through recommendations), the ramifications for our democracy are enormous. My research aims to address this problem by studying the phenomenon of algorithmically curated misinformation. At a high level, this research aims to develop auditing pipelines with computational human-AI workflows to identify, measure and defend against the misinformation surfaced by the algorithmic platforms. To gain an in-depth understanding of this complex phenomenon, I have divided my research into three synergistic phases, each phase addressing a unique aspect of the problem:

(1) **Phase1:** How can we design methodologies to audit and measure algorithmically curated misinformation on multiple online platforms across various user features (gender, age, political affiliation, etc.), user actions (click-action, like-action, follow-action, etc.), and high-impact events (elections, COVID-19, gun-shootings, etc.) for impactful and popular search queries?
(2) **Phase 2:** How can we develop novel human-AI workflows to scale misinformation annotations while taking a value sensitive design approach?
(3) **Phase 3:** How would users contest problematic algorithmic outputs? And how can search platforms design for contestability?

2 PHASE 1: DEVELOPING METHODOLOGIES TO AUDIT SEARCH SYSTEMS TO EMPIRICALLY MEASURE THE PREVALENCE OF ALGORITHMICALLY CURATED MISINFORMATION.

As part of Phase 1, I designed audit methodologies to measure the extent of algorithmically curated misinformation on multiple online platforms across various user features, user activities, and popular search queries. Using this methodology, I conducted an exhaustive set of carefully controlled experiments to audit social media search interfaces. Through the experiments, we investigated the role of personalization (due to user demographics, geolocation, account history, etc.) in amplifying misinformation. Below I briefly describe the audit experiments that I conducted on YouTube and Amazon platforms.

Completed work

Study 1: Auditing YouTube for perennial and demonstrably false conspiracy theories. In the study, we conducted audit experiments to investigate whether personalization (based on age, gender, geolocation, or watch history) contributes to amplifying misinformation. After shortlisting five popular topics known to contain misinformative content (Chemtrails, Flat Earth, Vaccine Controversies etc.) and compiling associated search queries representing them (via Google

CSCW '21 Companion, October 23–27, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8479-7/21/10.
https://doi.org/10.1145/3462204.3481788

Trends and YouTube auto-complete suggestions), we conducted two sets of audits—Search- and Watch-misinformative audits. Our audits resulted in a dataset of more than 56K videos compiled to link stance (whether promoting misinformation or not) with the personalization attribute audited. Our videos corresponded to three major YouTube components: search results, UpNext, and Top 5 recommendations. We found that demographics, such as, gender, age, and geolocation do not have a significant effect on amplifying misinformation in returned search results for users with brand new accounts. On the other hand, once a user develops a watch history, these attributes do affect the extent of misinformation recommended to them. **This work got published in CSCW 2020 [5].**

Study 2: Auditing Amazon for health misinformation. In this study, we conducted two-sets of algorithmic audits for vaccine misinformation on the search and recommendation algorithms of Amazon. First, we systematically audited search-results belonging to vaccine-related search-queries without logging into the platform—unpersonalized audits. Second, we analyzed the effects of personalization due to account-history, where history is built progressively by performing various real-world user-actions, such as clicking a product, adding product to cart, etc—personalized audits. Our work provides an elaborate understanding of how Amazon's algorithm is introducing misinformation bias in product selection stage and ranking of search results across five Amazon filters for ten impactful vaccine-related topics. Through our audit experiments, we also empirically establish how certain real-world actions on health misinformative products on Amazon could drive users into problematic echo chambers of health misinformation. **This work got published in CHI 2021 [6] and received a best paper honourable mention award.**

3 PHASE 2: DESIGNING HUMAN-AI WORKFLOWS FOR MISINFORMATION ANNOTATIONS USING A VALUE-SENSITIVE DESIGN APPROACH.

Having developed a methodology to audit and identify harms of AI-based algorithmic systems, Phase 2 of my research aims at designing and building human-AI workflows to annotate online content for misinformation at scale. The current automated fact-checking solutions fail to generalize to real-world fact-checking scenarios [4]. Furthermore, concerns have been raised automated systems using AI technologies might compromise important journalistic values such as transparency, accountability and responsibility [2, 7]. Thus, as a first step, we turned to fact-checkers to understand how they fact-check content belonging to different modalities (text, video, images, etc.). The next step is to use the insights provided by fact-checkers to develop to develop a tool to annotate online content collected in the audits for misinformation using the the capabilities of both humans and machines.

Completed work

Study 3: Understanding the needs,challenges and values of fact-checkers. Informed by value-sensitive design methodologies, we interviewed 18 fact-checkers from 11 fact-checking organizations, and identified important journalistic values in fact-checking as well as specific needs and challenges faced by the fact-checkers **(paper under review)**. The interviews revealed that algorithm explainability combined with tools that have humans in the loop emerged as key values that fact-checkers desire in the systems built for them. Fact checkers also expressed enthusiasm for tools to automate some of the manual verification process, for example, certain procedural tasks, like detecting style and language variations or computing metadata information such as the number of ads and extracting useful information from the comments section to get clues that would help them investigate the claims.

Proposed work

Study 4: Define workflows and develop systems for human-AI based fact-checking systems that upholds fact-checking values. Based on the interview insights, we propose an end-to-end automated system that will combine ML and NLP assisted workflows with fact-checkers'expertise to efficiently scale fact-checking without compromising quality and the values driving fact-checking practices. Our proposed method involves fact-checkers'knowledge and explicit feedback in all stages of tool development— requirement elicitation, feature engineering, system design, deployment and testing.

4 PHASE 3: DESIGNING TECHNOLOGIES TO CONTEST PROBLEMATIC ALGORITHMIC OUTPUTS

While audit studies conducted in Phase 1 provide a way to detect problematic behaviour in black-box algorithmic systems, it is also important to understand how users perceive such behaviour and how would they like to contest the problematic algorithmic outputs. Allowing users to contest the algorithmic decisions would help restore human agency in the algorithmic systems. How do people make sense of algorithmic output? When do people consider algorithmic output as problematic? What are the various ways in which people would like to contest the algorithmic output? How can search platforms design for contestability? These are some of the questions that the study conducted in Phase 3 of my research would answer.

Proposed work

Study 5: Design for algorithmic contestability. In this proposed research, first we will interview users of algorithmic systems to determine when do users consider algorithmic output as problematic. The outcome of the first part of the research would be a taxonomy of undesirable algorithmic outputs. Second, we will ask users what questions would they they like the platform to answer for each of the scenarios they mentioned in the first part (e.g. What, Why, Why not, When, etc.). Then, using the existing Explainable AI frameworks and curated questions, we will propose design ideas for contesting algorithmic platforms. We envision the use of such designs as 'decision aids' to users which will help them make informed choices on the platform. **An extended abstract of this work got accepted at CHI 2021 Workshop on Operationalizing Human-centered Perspectives in Explainable AI [9].**

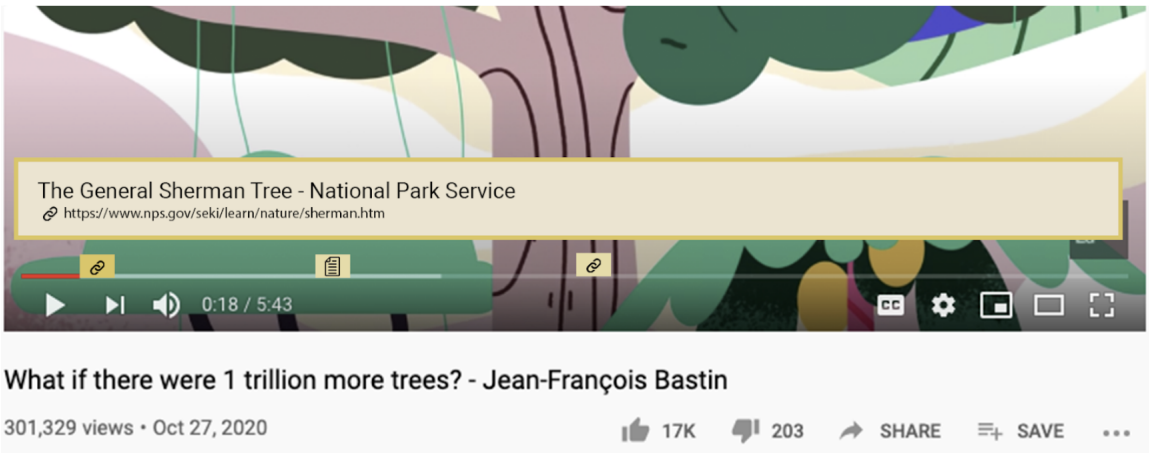

Figure 1: A design mock up illustrating how citations could look within a YouTube video page.

Study 6: Design for contesting the credibility of content surfaced by algorithms. In this study we explore several designs through which users can contest the credibility of content presented to them by algorithms. In particular, we explore design ideas to introduce citations in YouTube videos. Taking inspiration from how citations impart trust in Wikipedia articles [1], we explore the idea of how community of users can contest the content of videos presented to them by YouTube's algorithm by adding citation signals. **The initial work done for this study got accepted to ICWSM 2021 workshop on Information Credibility & Alternative Realities in Troubled Democracies [3].** See Figure 1 for a mock-up illustrating how citations could look on the video playback page on YouTube.

5 GOALS FOR CSCW DOCTORAL COLLOQUIUM

The CSCW Doctoral Colloquium would take place towards the end of the third year of my PhD. This would be the time when I'll be preparing for my PhD proposal. Therefore, the first goal in attending the colloquium is to gain feedback from the CSCW community about the framing of my research and determining whether the individual research studies that I conducted and have proposed together make a coherent thread. The colloquium would also be a great place to brainstorm alternative future research directions that my existing work could take. Finally, it would also provide me the opportunity to witness the doctoral work of my peers and engage with them in productive discussions.

REFERENCES

[1] [n.d.]. Research:The role of citations in how readers evaluate Wikipedia articles - Meta. https://meta.wikimedia.org/wiki/Research:The_role_of_citations_in_how_readers_evaluate_Wikipedia_articles. (Accessed on 06/10/2021).
[2] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.
[3] Prerna Juneja Tanushree Mitra Amy X. Zhang Emelia Hughes, Renee Wang. 2021. Introducing Credibility Signals and Citations to Video-Sharing Platforms. *ICWSM 2021 Workshop on Information Credibility & Alternative Realities in Troubled Democracies* (2021).
[4] D Graves. 2018. Understanding the promise and limits of automated fact-checking. (2018).
[5] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
[6] Prerna Juneja and Tanushree Mitra. 2021. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–27.
[7] Tomoko Komatsu, Marisela Gutierrez Lopez, Stephann Makri, Colin Porlezza, Glenda Cooper, Andrew MacFarlane, and Sondess Missaoui. 2020. AI should embody our values: Investigating journalistic values to inform AI technology design. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–13.
[8] BBC News. 2019. YouTube aids flat earth conspiracy theorists, research suggests. (2019). https://www.bbc.com/news/technology-47279253
[9] Tanushree Mitra Prerna Juneja. 2021. Algorithmic nudge: Using XAI frameworks to design interventions. *CHI 2021 Workshop on Operationalizing Human-centered Perspectives in Explainable AI* (2021).