Statement on Research

Today, we are surrounded by algorithmic systems in our everyday life. Although there are many cases where these systems are beneficial to users, scholars and regulators are concerned that they may also harm individuals and society. For example, sociologists and political scientists worry that online "filter bubbles" may create echo chambers that increase political polarization, while personalization on e-commerce sites can be used to implement price discrimination. Furthermore, algorithms may exhibit racial and gender discrimination if they are trained on biased datasets. As algorithmic system proliferate, the potential for (unintentional) harmful consequences to individual people and society as a whole increases.

The goal of my research is to examine the implications of opaque, pervasive algorithms in society. I am broadly focused on two complementary areas: the collection of data that becomes the *input* to systems and how the *output* of systems impacts people. I am also active in several related research areas, including online privacy, web security, and deceptive "dark pattern" user interfaces. My work is inherently interdisciplinary, drawing on computer science principles, as well as methods from sociology, law, economics, and political science. I owe a great debt to my many collaborators, including undergrads, Ph.D. students, faculty at Northeastern, and collaborators at a variety of universities.

In this statement, I present a brief overview of my research over the last five years since I received tenure. During this time period I published 33 papers in top-tier conferences (e.g., PETS, IMC, CSCW, CHI, FAccT, CCS, NDSS, WWW, ICWSM, SIGIR, ACL-IJCNLP) and journals (e.g., Nature, Science Advances, Harvard Kennedy School Misinformation Review), although I will only cover a subset in this statement. According to Google Scholar, as of May 10, 2023, I have 11,628 citations and an h-index of 49.[1]

I am proud that my work has had positive, real-world impact. We have helped companies fix serious bugs, several companies have changed their business practices in response to our studies, and our measurement studies are heavily cited in high-profile lawsuits. My work has been widely publicized in the media, which has helped raise awareness of algorithmic issues amongst the general public. Finally, I regularly work with government agencies to help them measure and regulate algorithmic systems.

My work is funded by the National Science Foundation (NSF), the Mozilla Foundation, the Knight Foundation, the Russell Sage Foundation, the Democracy Fund, the Anti Defamation League, Northwestern University, Underwriters Laboratories, the Data Transparency Lab, the European Commission, Google, Pymetrics, Verisign Labs, and by Northeastern University's TIER 1 program. I was named a Sloan Fellow in 2019. Post-tenure have been awarded grants and contracts totalling $17.5M and have received $75K in unrestricted gifts. This funding has enabled me to support five Ph.D. students and three undergraduate researchers. I have graduated four Ph.D. students in total (Le Chen, Muhammad Ahmad Bashir, Shan Jiang, and Ronald E. Robsertson), a fifth is defending in June 2023 (Avijit Ghosh), and a sixth proposed in May 2023 (Desheng Hu).

Algorithm Auditing

My work, along with my collaborators, is at the forefront of the growing area of *algorithm auditing*: we use carefully controlled experiments and observational data to understand the algorithms used by companies and assess their impact on normal people. Our ultimate goals are to make algorithmic systems more transparent and accountable to the public. We are actively collaborating with regulators to turn our research findings into policy outcomes.

Filter Bubbles. Since 2012 one thread of my research has been investigating the "filter bubble" effect on search engines. This theory argues that personalization of political search

[1] `https://scholar.google.com/citations?user=KzmHyt8AAAAJ`

results can result in a situation where partisans are only shown information that is congruent with their pre-existing beliefs, and that this may increase political polarization in the electorate. Two complementary papers authored with colleagues from the Network Science Institute paint a comprehensive picture that Google Search is not creating partisan filter bubbles. In the first study [6], published at CSCW in 2018, we recruited 187 people to take a survey and install a web browser extension, which we then used to execute queries for political topics chosen by us on Google Search. This method of data collection gave us the ability to observe the results Google showed to a wide spectrum of people as-if they were all searching for the exact same things at roughly the same time. In the second study [5], currently in press in Nature, we recruited 275 and 459 participants in 2018 and 2020, respectively, to take a survey and install a web browser extension, but in this case we simply observed and recorded the participants' Google Search queries, the search results they were shown, and their web browsing history. We analyzed these uncontrolled, observational behavioral traces and found that (after controlling for a variety of factors), Google Search was presenting the same mix of center-left content to all participants, regardless of their self-expressed political preferences. In contrast, our participants did tend to click on search results and visit webpages overall that were congruent with their political preferences.

Radicalization. Similar to the filter bubble, sociologists and journalists have argued that recommendation algorithms may be a mechanism for radicalization, with YouTube being singled out for particular criticism. In collaboration with political scientists from Dartmouth and University of Exeter, we studied whether video recommendations were actually radicalizing people [2]. In 2020 we recruited 1,181 participants to take a survey and install a web browser extension that recorded all of their activity on YouTube, including video recommendations they were shown. We found that, overwhelmingly, recommendations to videos from hyper-partisan and hate speech channels were shown to people who already viewed and subscribed to these channels. In contrast, within our six month observation window, only 30 participants followed a recommendation to a video from a problematic channels who were not already watching or subscribed to these channels. Although it may be the case that YouTube's algorithms recommended more problematic content to people before they made major changes in 2019, our results suggest that YouTube was not doing this at large-scale in 2020. This work is in press in Science Advances.

Content Moderation. Filter bubbles and radicalization rabbit holes are specific instances of concerns about content moderation on online platforms. There are more general concerns about content moderation espoused by political conservatives in which they claim that platforms are biased and unfairly censor their content. My Ph.D. student, Shan Jiang, and I debunked this claim using video comment moderation on YouTube as a case study. We repeatedly crawled comments on thousands of fact-checked YouTube videos to observe when comments were posted and removed. While we did observe more comment moderation on right-leaning videos, this effect was not statically significant once we controlled for the language in comments and video veracity. In other words, YouTube appeared to be faithfully applying their community guidelines that prohibit hate speech, strong language, and certain classes of misinformation—there just happens to be more comments fitting these criteria on conservative videos. This work was published in ICWSM 2019, where it received an Outstanding Analysis Award and was invited to appear at the main AAAI 2020 conference.

Bias and Discrimination. My algorithm auditing work extends beyond content moderation issues to other concerns, such as fairness and bias in machine learning (ML). In 2020, Alan Mislove, our Ph.D students, and I were approached by a startup called pymetrics who asked us to audit their product—an ML-based screening tool for job applicants—for racial and gender bias. This was a challenging project because it was, and remains to this day, one of the only audits ever conducted "cooperatively" between external experts and a private company. We designed careful contractual and interpersonal protocols to preserve our independence as auditors, including shielding our testing methods from pymetrics so they could not "game" our tests. While we found some minor areas of concern, overall pymetrics passed our audit. We (the Northeastern team and pymetrics) jointly published a paper detailing the audit at FAccT 2021 [7], with the goal being to promote more audits of this kind by other teams and companies in the future.

My work on algorithm auditing is funded by a NSF CAREER award, a NSF Small award, and grants from the Sloan Foundation, the Russell Sage Foundation, the Anti-Defamation League, and Pymetrics totaling $1.4M.

The National Internet Observatory

Along with my co-PIs David Choffnes and David Lazer, we have launched a new initiative called the National Internet Observatory (NIO) that aims to make algorithm auditing much more accessible to the research community. With $15.7M in NSF support, we are recruiting a large-scale, permanent pool of participants who will be regularly surveyed and will consent to provide data about their online activity on the web and mobile devices. Our mandate is to make all of this data available to the research community, while ensuring the highest standards of ethical data collection and participant privacy protection. We hope that the NIO will revolutionize the study of online life in general, as well as foster a community of algorithm auditors who will investigate major platforms, make their practices more transparent, and increase accountability in the tech space. As of May 2023 the NIO is in the field with around 400 participants, and we are planning to ramp up subject recruitment while also welcoming our first cohort of researchers in Fall 2023.

Online Privacy

Closely related to my algorithm auditing work is another thread of research focused on privacy. Whereas algorithm audits focus on the outputs of systems, this work is interested in the inputs, i.e., the personal data that is harvested about users by online services. One example of my research in this area is a study that I conducted with my Ph.D. student, Muhammad Ahmad Bashir, and collaborates from LUMS in Pakistan in which we examined the accuracy of demographic and behavioral profiles inferred by online advertisers [1]. These "targeting profiles" are the central value-proposition of online advertising—advertising platforms claim that they allow advertisers to direct ads to people with unparalleled accuracy and precision, thus justifying the high costs of online ads. We recruited 220 people to install a web browser extension that allowed us to (1) record copies of the targeting profiles Google and Facebook had built about participants and (2) ask participants whether the contents of the profiles were accurate. We found that well over half of the inferences were inaccurate and participants said that ads targeted to those attributes would not be useful to them. This paper was published at NDSS 2019, and the results challenge the very foundations of the online advertising economy.

The online privacy landscape is changing rapidly due to the passage of new, comprehensive privacy laws like GDPR. However, it remains unclear whether these laws are having their intended effects. Northeastern undergrad Maggie Van Nortwick and I conducted a study, published at PETS 2022 [4], in which we set out to measure websites compliance with the California Consumer Privacy Act (CCPA), the strongest online privacy law in the US. Unlike GDPR, the CCPA does not apply to all websites, so a key facet of our study was determining which websites met the CCPA's eligibility criteria by estimating the number of unique visitors they had from California. Overall, we found that compliance with the most basic facets of the CCPA were low, even when accounting for the law's eligibility criteria. These results demonstrate who strong enforcement of privacy laws is critical and lay bare the long road ahead for regulators. Maggie's work was funded by an NSF REU and she received an Undergraduate Research Award from Northeastern for this study.

Cybersecurity

In addition to my work on algorithm auditing and online privacy, I am actively researching traditional areas of cybersecurity: specifically, I am part of a long-running collaboration between researchers from Northeastern, Virginia Tech, Duke, Carnegie Mellon, and U. of Maryland that has been measuring and devleoping novel systems to improve Public Key Infrastructures (PKIs).[2]

The most recent product of this collaboration is a system called Hammurabi that enables TLS clients (e.g., web browsers, command line tools, libraries like OpenSSL, etc.) to separate X.509 certificate validation policy (e.g., minimum key sizes, maximum certificate lifetimes, etc.) from mechanism [3]. Our prototype uses Prolog-based logic programs to specify certificate validation policies. This separation allows TLS clients to specify concise policies (1̃00 lines of code) that are cleanly delineated from the thousands of lines of (typically C) code that parse X.509 and implement cryptographic primitives, rapidly adopt strong policies authored by trusted parties, and even impute the functional differences between policies. We reimplemented Firefox and Chrome's validation policies in Prolog and integrated Hammurabi into Firefox and Golang to demonstrate the benefits of our approach. Hammurabi was published at CCS 2022 and received a Honorable Mention Award.

My work on the PKI and TLS is funded by an NSF Medium award of $599K ($1.2M total) that is a collaborative grant with PIs at the U. of Maryland, and an NSF Large award of $400K ($1.2M total) that is a collaborative grant with PIs at Virginia Tech, U. of Maryland, CMU, and Duke.

References

[1] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. Quantity Vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers. In *Network and Distributed System Security Symposium*, San Diego, CA, 2019.

[2] Annie Y. Chen, Brendan Nyhan, Jason Reifler, Ronald E. Robertson, and Christo Wilson. Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. *In Press: Science Advances*, 2023.

[3] James Larisch, Waqar Aqeel, Michael Lum, Yaelle Goldschlag, Leah Kannan, Kasra Torshizi, Yujie Wang, Taejoong Chung, Dave Levin, Bruce M. Maggs, Alan Mislove, Bryan Parno, and Christo Wilson. Hammurabi: A Framework for Pluggable, Logic-based X.509 Certificate Validation Policies. In *ACM Conference on Computer and Communications Security (CCS 2022)*, Los Angeles, CA, 2022.

[4] Maggie Van Nortwick and Christo Wilson. Setting the Bar Low: Are Websites Complying With the Minimum Requirements of the CCPA? *Proceedings on Privacy Enhancing Technologies (PoPETS)*, 2022(1), 2022.

[5] Ronald E. Robertson, Jon Green, Damian J. Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer. User choice outweighs algorithmic curation for partisan news on Google Seach. *In Press: Nature*, 2023.

[6] Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. Auditing Partisan Audience Bias Within Google Search. *Proceedings of the ACM: Human-Computer Interaction*, 2(CSCW), 2018.

[7] Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, and Frida Polli. Building And Auditing Fair Algorithms: A Case Study In Candidate Screening. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT 2021)*, Virtual Event, Canada, 2021.

[2] More information is available at `https://securepki.org/`.