Motahhare Eslami Research Statement

My research endeavors to understand and improve the interaction between users and opaque algorithmic socio-technical systems. Algorithms play a vital role in curating online information in socio-technical systems, however, they are usually housed in black-boxes that limit users' understanding of how an algorithmic decision is made. While this opacity partly stems from protecting intellectual property and preventing malicious users from gaming the system, it is also designed to provide users with seamless, effortless system interactions. However, this opacity can result in misinformed behavior among users, particularly when there is no clear feedback mechanism for users to understand the effects of their own actions on an algorithmic system. The increasing prevalence and power of these opaque algorithms coupled with their sometimes biased and discriminatory decisions raises questions about how knowledgeable users are and should be about the existence, operation and possible impacts of these algorithms.

My work draws on human-computer interaction, social computing and data mining techniques to investigate users' behavior around opaque algorithmic systems and create new designs that communicate opaque algorithmic processes to users and provide them with a more informed, satisfying, and engaging interaction. In doing so, I add new angles to the old idea of understanding the interaction between users and automation by 1) investigating algorithmic effects on users' experience, 2) designing around algorithm sensemaking, 3) designing for algorithmic transparency, and 4) auditing and designing around algorithmic bias.

ALGORITHMIC EFFECTS ON USER'S EXPERIENCE

To evaluate how algorithms shape and influence user's experience, I have investigated users' behavior around algorithms that followed the same goal but generated different outputs. I developed an application, *GroupMe*, that applied three different clustering algorithms on a user's Facebook friendship network (Figure 1) [1]. These algorithms used this network as the input to create groups of friends automatically, but via different methods which resulted in different groupings.

Figure 1. *GroupMe* Facebook Application

To examine how a grouping algorithm impacts a user's perception of and interaction with her friendship groups, I asked some Facebook users to use *GroupMe* to modify the generated grouping by each algorithm and create their final "desired groupings". This process resulted in three desired groupings for each user. I then measured the similarity between each pair of desired groupings created by each user and found a 14% difference on average between a user's final desired groupings. Patterns of use and interview results showed the reason behind this major difference was *following what algorithms create*: users stated that if an algorithm did not find a specific group, they might have not created it themselves, but when a group was created, they usually liked it, and kept it. This shows that the choice of using a different algorithm can shape a users' experience differently.

DESIGNING AROUND ALGORITHM SENSEMAKING

Observing the great power of algorithms in shaping users' experience raised questions about how aware users are of such algorithmic impacts, and what factors impact this awareness [2]. To answer these questions, in a series of interviews with Facebook users, I investigated users' awareness of the Facebook News Feed curation and found that the majority of users were not aware that their feed was filtered algorithmically [3]. Qualitative and quantitative analysis of users' usage behavior showed that this lack of awareness was related to the level of users' engagement with their feed: the less users were actively engaged with their feed, the less they were aware of their feed algorithmic curation.

To increase users' awareness and sensemaking of their algorithmic feed curation process, I developed a Facebook application, *FeedVis*, that incorporated some "*seams*", visible hints disclosing aspects of automation operations, to the opaque feed curation algorithm. *Feedvis* discloses what I call "*the algorithm outputs*": the differences in users' News Feeds when they have been curated by the algorithm and when they have not (Figure 2). Walking through this seamful design, users were most upset when close friends and family were not shown in their feeds. I also found users often attributed missing stories to their friends' decisions to exclude them rather

than to Facebook News Feed algorithm. By the end of the study, however, users were mostly satisfied with the content on their feeds. A follow up with the users showed that algorithmic awareness led users to more active engagement with Facebook and bolstered overall feelings of control on the site.

(a) (b)

Figure 2. (a) *FeedVis Content View* highlights the content that the algorithm excluded from display. (b) *FeedVis Friend View* reveals social patterns by disclosing whose stories appeared and whose were hidden in News Feed.

This work was awarded Best Paper at CHI 2015 and has been cited over 160 times since it was published. It has been also discussed in many popular press, including Time, The Washington Post, Huffington Post, the BBC, CBC Radio, Fortune, Quartz, International Business Times, New Scientist, and MIT Technology Review. The work has also been included in the curricula of several graduate human-computer interaction courses, including at Carnegie Mellon University and the Max Planck Institute.

Folk Theories of Algorithm Operation. To understand the impacts of a seamful design on users' perceptions of an opaque algorithm, I discovered "folk theories" that users developed about how the algorithm works before and after being exposed to some hidden aspects of the algorithm via *Feedvis* [4]. Patterns of use and results showed that incorporating intentional seams into the feed helped users who were unaware of the algorithm's existence develop theories similar to users who were aware of the algorithm's presence. This rapid similarity suggests that providing extra visibility into an algorithm could help users rapidly develop new and predictable conceptual understandings of an algorithmic system.

DESIGNING FOR ALGORITHMIC TRANSPARENCY

So far, my research has shown that disclosing aspects of an algorithmic process could shape an informed and engaging interaction between users and the system. But how much transparency is enough or even practical in opaque algorithmic systems? To answer this, I studied users' perception of and attitudes towards different transparency mechanisms in online behavioral advertising—I have particularly chosen this domain due to the opaqueness of personalization advertising algorithms and users' privacy concerns about them [5]. I exposed users to a) why a specific ad is shown to them, b) what attributes an advertising algorithm infers about them, and c) how an advertiser uses this information to target users. The results showed that either vague and oversimplified language or very specific explanations about how users' online ads were tailored to them were unsatisfying. Users were most satisfied with ad explanations that included *some* specific information that an advertiser used to target an ad, particularly if it was an important and recognizable part of user's identity.

Algorithm Disillusionment. I also learned that disclosing algorithmically-inferred attributes and interests, particularly those that are wrongly inferred, can lead users who assumed algorithmic authority—that advertising algorithms are perceptive, powerful, and sometimes scary—to algorithm disillusionment—that algorithms are not scary and powerful, or even effective. This realization can help users not to trust algorithms unconditionally, and therefore have a more realistic interaction with algorithmic systems.

AUDITING & DESIGNING AROUND ALGORITHMIC BIAS

When an opaque algorithm is biased or is suspected to be biased, I take further steps to build a more informed interaction between users and such algorithms: a) First, I develop audit techniques to detect and quantify algorithmic bias, b) I then explore users' understanding of and behavior around detected biases, and c) finally, I use this information to build a design that adds transparency into a biased algorithm to investigate the impacts of transparency on users' attitudes and intentions.

A) Algorithm Auditing: Detecting and Quantifying Algorithmic Bias

To detect and quantify potential biases in black-boxed algorithmic systems, I developed different "cross-platform audit" techniques which determined whether an algorithm introduced bias to a system by comparing that algorithm's outputs with other algorithms' outputs of similar intent [6,7,8,9,10]. I particularly audited two categories of algorithmic systems which their opacity along with their power have raised concerns about the bias they might introduce into users' experience: search engines, rating platforms, and online housing.

Search Engines: In collaboration with my colleagues at the Max Planck Institute, we quantified and compared the political bias that Google and Twitter search can introduce to users' search results about 2016 U.S. presidential candidates [6,7]. This analysis showed that while the political bias of search results for a candidate name on Google was toward that candidate's party leaning, the political bias of search results on Twitter Search, regardless of the candidate's political leaning, was mostly favoring the democratic party. We found that a part of this significant difference came from the fact that in Google, a large fraction – 40.6% on average – of the results for the presidential candidates are from sources they control, i.e., either their personal websites or their social media profile links; this fraction, however, is much smaller for most candidates on Twitter – only 7.25%. In addition, our analysis showed that the full tweet stream containing political query-terms that build the input data to the search algorithm on Twitter contains a democratic slant and the algorithm usually strengthens this bias. This calls for new design approaches to increase users' awareness of such potential biases, and that their choice of the search engine can affect their political view.

Rating Platforms: In two other auditing efforts, I found that the opacity of online rating platforms in how their rating algorithms calculate a business's final rating can introduce bias to user's experience. In *Booking.com*, a hotel rating platform, a misrepresentation in the lowest possible review score allowed its rating algorithm to bias ratings of low-to-medium quality hotels up to 37% higher than three other hotel rating platforms (Expedia.com, Hotels.com, and HotelsCombined.com) [8]. In *Yelp.com*, a business rating platform, I found that its interface misrepresents whether a user's review is filtered or not. That is, Yelp only reveals that a user's review is filtered when the user is logged out. When logged in, the user sees her filtered reviews under the recommended reviews of a business (as if unfiltered). So a user can only detect if their reviews are filtered by looking for their own reviews when logged out or logged in as another user. I call this a bias since Yelp deceives a user by telling her that her review is not filtered, while it actually is [9].

Online Housing: We also designed an online housing auditing infrastructure by employing a sock-audit technique to build online profiles associated with a specific demographic profile. We used this infrastructure to examine if online housing ads as well as online housing listings exhibit discrimination against protected features like race and gender [10].

B) Users' Behavior around Algorithmic Bias

When an algorithmic bias is detected, I design studies to explore how users behave around such biases in order to build more informed interaction between users and biased algorithmic systems.

In the case of Booking.com, I first applied a computational technique to identify the users who noticed the bias and discussed it in their reviews [8]. The analysis of these discussions showed that detecting the bias made these users deviate from contributing the usual review content (i.e., informing other users about their hotel stay experience) and rather adopted a "*collective auditing*" practice: when users confronted a higher than intended review score, they utilized their review to raise the bias awareness of other users on the site. They wrote about how they: engaged in activities such as trying to manipulate the algorithm's inputs to look into its black-box, tried to correct the bias manually, and illustrated a breakdown of trust.

In another technique, I utilized the online discussion posts on the Yelp forum about the Yelp review filtering algorithm, along with interviews, to understand Yelp users' perceptions of and attitudes towards this algorithm, and its bias in both existence and operation [9]. The results showed that users took stances with respect to the

algorithm; while many users challenge the algorithm, its opacity and bias, others *defend* it. I found that the stance the user takes depends on both their personal engagement with the system as well as their potential of personal gain from the algorithm's presence.

C) Adding Transparency into a Biased Algorithm

When I understood users' behavior around the biased Yelp filtering algorithm, I developed *ReVeal* (Review Revealer), a tool that discloses the algorithm's existence by showing users which of their reviews the algorithm filtered (Figure 3) [9]. When evaluating the tool and discovering their filtered reviews, some users reported their intention to leave the system, as they found the system deceptive. Other users, however, report their intention to *write for the algorithm* in future reviews; i.e. they described some of the folk theories they developed during the study about how the algorithm works and stated that they would apply these theories to their future reviews to avoid their reviews being filtered. This shows that adding transparency to biased algorithmic systems allow users to have a more informed and adaptive interaction with the system to achieve their goals.

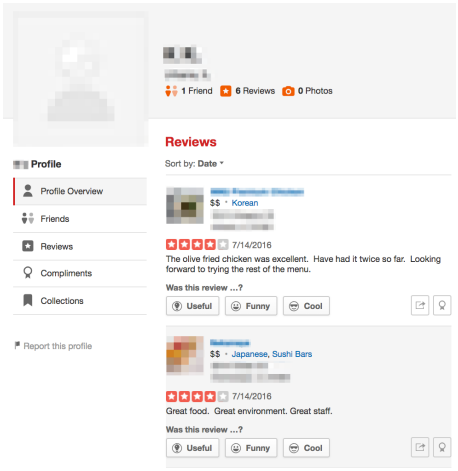

Figure 3. *ReVeal:* The tool shows users both their filtered and unfiltered reviews. Filtered reviews are highlighted with a gray background.

RESEARCH AGENDA

My goal is to improve users' interaction with automation in the era of AI and opaque algorithms. *GroupMe*, *FeedVis* and *ReVeal* are typical of my approach: I pick an algorithmic system which its algorithm's opacity might result in misinformed behavior among users, understand users' behavior around that algorithm, build designs that add transparency to the system, and investigate the impacts of the added transparency on users' behavior. To achieve my goal, I outline some future opportunities that I am excited to pursue.

The Future of Algorithmic Transparency. While algorithmic transparency was started as a topic of interest among researchers, it has now been considered by many other groups like activists, regulators, and even governments. One example is the European Union's new General Data Protection Regulation and providing users with a "right to explanation" about algorithmic decisions that were made about them. Such transitions in algorithmic systems, however, are not straightforward. While transparency might seem simply to help users understand algorithmic decisions better, it can also be detrimental: The wrong level of transparency can burden and confuse users, complicating their interaction with the system. Too much transparency can also disclose trade secrets or provide gaming opportunities for malicious users. These challenges have motivated me to explore different levels of transparency in algorithmic systems, particularly those that their decisions significantly affect users. For example, I am currently collaborating with the advertising team in Adobe Research to add various types of explanations in users' real ads and track their usage behavior in the wild. My goal is to understand what level of transparency provides users with a more informed interaction with their ads. I hope this could be a starting point for tackling this complex, and critical challenge.

Users as Auditors: **Algorithm Bug Bounty.** Amidst the numerous proposals to better understand opaque algorithmic systems, one thrust has focused on auditing these systems. These methods, from studying the code directly to collaborative audits, all require the intervention of researchers, regulators or other third parties to coordinate. My research on biased algorithmic systems, however, highlights a new form of audit: a *collective audit*, driven purely by users in a collective attempt to detect and understand algorithmic bias (such as in Booking.com and Yelp.com). This audit technique provides a "watchdog from within" practice. Looking for bias from the viewpoint of regular use increases the likelihood of detecting bias, as well as the likelihood of other users becoming aware of the bias. We, however, lack mechanisms that enable collective audit efforts in a systematic and organized manner among users. I am currently collaborating with researchers from Northeastern University and University of California, Berkeley to investigate the design practices that could support users reporting biases. One of these practices which is used in the security area is "*bug bounty*" programs: companies incentivize system users to conduct security research and report flaws for monetary and reputational gain while providing legal protection from the applicable anti-hacking laws. Transferring such practices to the algorithm bias domain can develop an ecosystem empowering users, and foster auditing broadly.

When Human Bias and Algorithm Bias Collude. While algorithms sometimes introduce bias and discrimination into a user's experience, they are not the only party to blame. In many cases, algorithmic bias spreads to a system through a training dataset from biased individuals. In other cases, human bias itself reinforces algorithmic bias significantly. For example, while ideological filter bubbles can occur by algorithms that filter the content a user might not like, human's desire for selective exposure (people's preference to view content they agree with to get more self-assurance) can be as powerful in creating or reinforcing filter bubbles. I am currently exploring political filter bubbles as a type of bias that both human and algorithms have a role in creating it. While so far my research has focused on the algorithmic side of a bias, I have a long-standing interest in understanding the dynamics of systems that both human bias and algorithm bias are involved in: How do these two types of bias interact? And what are the ways to detect, distinguish and mitigate these biases?

Moving Forward. None of the above goals, however, are possible without collaborating with and learning from researchers from different disciplines and backgrounds. In graduate school, I have been fortunate to collaborate with many researchers from areas of computer science (human-computer interaction, data mining, and artificial intelligence), information and communication studies, and art and design. My collaborators come from more than ten academic departments and research labs, and I hope that I can continue and expand this tradition of collaboration as I move forward.

In the long term, my research interests are framed by what I identify as real-world challenges, from users' misinformed usage of their social media feeds or search engines, to privacy challenges algorithms cause by misusing users' private information in targeting ads to them, to deceiving users to book a hotel with a low quality. To these challenges I bring a technical approach and an understanding of computer science and human-centered design.

REFERENCES

[1] **M. Eslami**, A. Aleyasen, R. Zilouchian Moghadam and K. Karahalios. *Friend Grouping Algorithms for Online Social Networks: preference, bias, and implications*. The 6th International Conference on Social Informatics (SocInfo), 2014.

[2] K. Hamilton, K. Karahalios, C. Sandvig, and **M. Eslami**. *A Path to Understanding the Effects of Algorithm Awareness.* The Human Factors in Computing Systems Conference *(CHI),* Alt.CHI, 2014.

[3] **M. Eslami**, A. Rickman, K. Vaccaro, A. Aleyasen, A. Vuong, K. Karahalios, K. Hamilton, and C. Sandvig. *"I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in the news feed. The Human Factors in Computing Systems Conference (CHI), 2015.* ***Best Paper Award.***

[4] **M. Eslami**, K. Karahalios, C. Sandvig, K. Vaccaro, A. Rickman, K. Hamilton, and A. Kirlik. *First I "like" it, then I hide it: Folk Theories of Social Feeds*. The Human Factors in Computing Systems Conference (CHI), 2016.

[5] **M. Eslami**, S. R. Krishna Kumaran, C. Sandvig, and K. Karahalios. *Communicating Algorithmic Process in Online Behavioral Advertising*. The Human Factors in Computing Systems Conference (CHI), 2018.

[6] J. Kulshrestha, **M. Eslami**, J. Messias, M. B. Zafar, S. Ghosh, K. Gummadi, and K. Karahalios. *Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media*. The Computer-Supported Cooperative Work and Social Computing Conference (CSCW), 2017.

[7] J. Kulshrestha, **M. Eslami**, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K Karahalios. *Search Bias Quantification: Investigating Political Bias in Social Media and Web Search*. Information Retrieval Journal, 1-40, 2018.

[8] **M. Eslami**, K. Vaccaro, K. Karahalios, and K. Hamilton. *"Be careful; things can be worse than they appear": Understanding Biased Algorithms and Users' Behavior around Them in Rating Platforms*. The International AAAI Conference on Web and Social Media (ICWSM), 2017.

[9] **M. Eslami**, K. Vaccaro, M. K. Lee, A. Elazari, E. Gilbert, and K. Karahalios. *User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms*. The Human Factors in Computing Systems Conference (CHI), 2019.

[10] J. Asplund, **M. Eslami**, R. Barber, H. Sandaram, and K. Karahalios. *Auditing Race and Gender Discrimination in Online Housing*. *Submitted* to The International AAAI Conference on Web and Social Media (ICWSM), 2019.