Rakibul Hasan Research Statement

Overview. I aim to minimize the privacy/security risks of existing and emerging technologies, while preserving their utility and ensuring equal benefits for all. My doctoral research was focused on social networking platforms. My future research will encompass wearable devices, smart home automation and surveillance systems, and educational technologies, such as remote proctoring tools. In my research, I take a holistic approach—combining technical, psychological, and social means—to both gain a deeper understanding of the privacy/security threats posed by such technologies and mitigate them using machine learning, computer vision, human-computer interaction, cognitive psychology, and social science. I conduct research through controlled experiments [1-6] as well as by collecting large scale observational data over multiple years [7].

Impact of my doctoral research. My efforts (in collaboration with my Ph.D. advisor and experts in psychology) to understand people's online photo-sharing behaviors were awarded an **NSF SaTC (Secure and Trustworthy Cyberspace) grant**. My research has been published in prominent security/privacy and HCI conferences (e.g., **IEEE S&P** [1,5], **CHI** [2,3,6]) as well as awarded by internationally recognized organizations (e.g., **INRIA**).

My research philosophy and approach.

I aim to Build *transparent* and *trustworthy* technologies. I strive to ensure the highest level of transparency and explainability while developing privacy-enhancing technologies (PETs). For example, while building machine learning models to classify people in images in [1], first, we inferred a few *high-level*, *intuitive* features using very high-dimensional features obtained from deep learning models. Those inferred features—*which we designed to replicate human reasoning processes that we measured through a user study*—were used to train the final classifier. Thus, decisions made by these models can be easily explained, which **provides transparency and helps to enhance trust** in the system.

Privacy/security enhancing technologies are not useful if they are not *usable*. To be usable and widely accepted, security/privacy-enhancing features of any technology should preserve the technology's utility and minimally disrupt its workflow. While designing and evaluating image obfuscations [2, 3], I considered both privacy-protection capability and utility (visual aesthetics and viewers' satisfaction) that are very important to photo-sharers for *self-presentation* and *eliciting social interactions*, as identified by prior research. Further, we proposed artistic image and video transforms to improve the visual aesthetics while enhancing privacy [3, 4].

I focus on both advancing scientific knowledge and having practical impact. My research is driven by the aspiration to create new knowledge and solve practical problems using it. A central goal of my past and ongoing research has been to understand people's decision-making process in the context of privacy-sensitive information sharing [6], identifying internal and external factors that influence their behaviors [6, 7], and how can they be encouraged in privacy-preserving behaviors ([5, 6]). I have analyzed large-scale data from the wild using a

causal inference framework [7] to discover the underlying reasons behind changing people's online behavior that would provide actionable insights, instead of correlational analyses that reveal only associations. In [1-4], I built and evaluated privacy-enhancing technologies that are effective, practical, and usable.

I approach problems in a multidisciplinary nature. Computing technologies have penetrated every aspect of our lives, and thus, they demand multidisciplinary efforts to build and study them. I have active collaborations with experts in *security* and *privacy*, *machine learning and computer vision*, *human-computer interaction*, *cognitive psychology*, *social science*, and *economics*. As I broaden my research vision and scope post-graduation, I will actively expand and diversify the network of collaborators, in particular, to include experts in *complex and dynamic systems analysis*, *internet measurement*, *systems security*, and *policy*.

Future research plans

Making online social media safer. With just a click, people can now share a piece of information with the whole world, and that may result in severe consequences for themselves or other people. Online social networks are driven by their users; they simultaneously create, propagate, and consume content on these platforms. Thus, I strongly believe that the users' active participation is essential in alleviating the privacy, safety, and security threats in this connected space. Through a concerted effort in collaboration with experts in cognitive psychology, sociology, HCI, and machine learning, I will build usable, intelligent, and adaptive privacy-enhancing tools, as well as design behavioral interventions that would encourage privacy-preserving, respectful, and prosocial behaviors by using those tools.

Thwarting the dissemination of misinformation through online platforms. Rumors and misinformation play a key role in shaping public opinion, which in turn influences laws and policies at the national and global levels that affect the lives of millions. I will contribute to obstructing the proliferation of misinformation by i) modeling human behaviors in the context of *believing* and *propagating* misinformation and identifying the *causal mechanisms* behind those behaviors, ii) designing and evaluating interventions to inspire critical thinking and better judgement, and iii) identifying mechanisms and patterns of how misinformation spreads across communities so that it can be spotted early and prevented from spreading more widely. Toward this goal, I plan to conduct large-scale empirical research in collaboration with scholars in internet measurement, complex network analysis, social science, data mining, and pattern recognition.

Identifying and mitigating privacy issues of smart home and surveillance systems. Technologies for home automation and surveillance, which are often criticized for questionable data collection and sharing practices, are becoming commonplace. In this context, I will research leveraging the power of statistical signal processing, machine learning, and computer vision techniques to detect and selectively obfuscate information (e.g., by introducing carefully designed noise) that is not essential to performing the primary functionalities of these systems but can potentially harm the owners' or other stakeholders' privacy. Furthermore, as forced by the recent pandemic, personal living spaces are increasingly proxying for workplaces. This is a very interesting setting in which to study people's adoption of home automation and surveillance systems, as well as their attitudes and concerns regarding information sharing with those systems in different personal, professional, and mixed contexts.

Discovering and alleviating the harmful impacts of *digitizing and datafying* educational processes. This is the advent of an era where an enormous amount of data about students is being gathered and analyzed, with a vision to shape their future prospects based on their *'performance' as students*. Educational institutes are increasingly deploying technologies to continuously monitor students and measure their engagement and progress, with the promise to provide need-based, personalized guidance and assistance. This vast amount of information is shared among many business entities and their 'partners,' who develop, deploy, and maintain them. Such opaque data-sharing practices pose significant threats to students' privacy and security. Another pernicious, but potentially more damaging consequence of these technologies might be that students will be implicitly encouraged to self-censor their behaviors as they are continuously monitored, profiled, ranked and compared with others, and sometimes flagged (e.g., when the system detects 'anomalous' behavior), which would hurt their independence and autonomy. Moreover, such technologies attempt to detect 'anomalous' behaviors by learning what is 'normal' from data that are often biased against minority groups and people with disabilities, thus disproportionately affecting them, as they look or act differently than the 'norm.' My research effort will be directed toward unveiling and quantifying the threats they pose toward students' privacy, security, safety, and mental well-being, as well as inventing mechanisms to mitigate them.

Ensuring the privacy and fairness of remote tutoring and proctoring tools. The recent pandemic has catalyzed a global surge in the adoption of remote tutoring and proctoring tools. They have great potential to democratize education by expanding the scale and driving down the expense. However, there are significant privacy risks to the students (and others living in the same space, e.g., family members and roommates) as these tools collect continuous audio and video data in home contexts. Students coming from lower socioeconomic backgrounds or densely populated regions will be disproportionately affected by the privacy risks. I will study how computer vision and machine learning-based technologies can help alleviate these risks, e.g., by selectively obfuscating the audio or video data to obscure sensitive information. Remote tutoring and proctoring tools employ artificial intelligence to automatically detect *unintended behaviors* (e.g., 'inattentiveness' during a lecture or 'cheating' during an exam), leaving ample room for discrimination against students whose outlooks or behaviors deviate from the 'norm' (e.g., showing unique facial expressions, having physical limitations such as low vision, or needing to visit the restroom more frequently than 'usual'). My research will attempt to improve these tools using statistical and machine learning models and empirically validate them in naturalistic settings using large scale cross-cultural studies, both *observational* and *experimental*.

Past and ongoing research

Problem space. My doctoral research goal was to *mitigate privacy risks in the context of photo-sharing on social media*. I emphasized both on photo-sharers and people whose privacy could be violated when *other* people share their photos—e.g., `bystanders' (i.e., people who get captured by chance in others' photos) and people whose photos have been used to create memes.

Automatically detecting bystanders in images to protect their privacy. We proposed a *fully automated* system to protect bystanders' (i.e., strangers who were captured in the images by chance) privacy [1], thus establishing the norm of **by-default-private**. We developed a general-purpose, machine learning-based model, which can be deployed on any photo-hosting platform and can be applied over all past, present, and future images, thus enabling privacy-protection at scale. To ensure that our model was transparent, we trained it using a few high-level, intuitive features that were identified based on how humans distinguish bystanders from photo subjects. This approach facilitates understanding and explaining the rationale behind classifying a person as a bystander by examining feature values, which is not possible if high-dimensional features (e.g., from deep learning models) are directly used to train the model.

Obfuscating privacy-sensitive content in images. The goal of this project was to design and evaluate image obfuscation methods in terms of their capability to protect privacy and preserve utility (e.g., visual aesthetics) for viewers. We applied eleven commonly used image filters on twenty privacy-sensitive objects and assessed their usability through a user study (N=570) [2]. We found that they failed to provide adequate privacy protection in most of the cases or degraded the image quality to the extent that the result was not acceptable to the viewers. In a later study [3], we closely examined the interactions among three utility variables (information content, visual aesthetic, and viewers' satisfaction) using path model analysis. The findings shed light on how each of them influences the others and offered a principled way to design novel obfuscations to balance privacy-utility trade-offs. Following this, we designed and evaluated new obfuscations by combining stylistic image transforms with privacy filters (N=653) [3].

Understanding people's online photo-sharing decisions and designing interventions to encourage privacy-respecting behaviors. In this project, we aimed to understand what factors (e.g., photo contents, individuals' personality traits, and social context) influence online image-sharing. Our ultimate goal is to design behavioral interventions to discourage the sharing of privacy-sensitive photos and encourage the adoption of privacy-enhancing technologies (e.g., to detect and obfuscate sensitive content in an image before sharing it). We started with simple privacy nudges (e.g., warnings about privacy risks), but obtained very surprising results—study participants (N=444) actually *increased* photo sharing after being primed [5]. Digging deeper, we found that many complex interactions among various personal and contextual factors are at play when people decide to share photos, hinting that *seemingly obvious* solutions may worsen the problem they were invented to solve [5]. To better understand photo-sharing behaviors, we investigated how personality traits such as humor styles (based on the work of Martin *et al.*) affect people's sharing preferences and reactions to behavioral interventions through a user study (N=453) [6]. Based on our experimental data, we discovered that people with certain 'humor types' are more likely to violate others' privacy, even after being explicitly warned. Currently, I am investigating how attention, cognitive load, and emotional arousal dictate photo-sharing behaviors using eye-tracking, cursor movement, and physiological data (such as pupil dilation and heart rate).

Quantifying how external stimuli affect people's online behaviors. In [7], I collected tweets, profile information, and the changes in follower graphs of more than 17,000 Twitter users for more than three years. I analyzed this longitudinal data in a **causal inference framework** to capture complex network dynamics. Our primary interest was to discover how online activities (e.g., posting tweets) change as a response to external stimuli (e.g., going 'viral' or suddenly accumulating a large number of followers).

Impact of attention, emotion, and cognitive load in information-disclosing behaviors. I am co-leading a project where we have conducted lab-based experiments and collected eye-tracking, cursor movement, heart rate, and pupil dilation data while participants were engaged in sharing potentially privacy-violating information. We are analyzing these data to understand how attention, emotional arousal, and cognitive load influence people's decisions to disclose information.

References

[1] Rakibul Hasan, David Crandall, and Mario Fritz Apu Kapadia. Automatically Detecting Bystanders in Photos to Reduce Privacy Risks. In 2020 IEEE Symposium on Security & Privacy (**Oakland'20**).

[2] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In Pro- ceedings of the 2018 CHI Conference on Human Factors in Computing Systems (**CHI '18**).

[3] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. Can privacy be satisfying? On improving viewer satisfaction for privacy-enhanced pho- tos using aesthetic transforms. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (**CHI'19**).

[4] E T Hassan, Rakibul Hasan, P Shaffer, D Crandall, and A Kapadia. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (**CVPRW'17**).

[5] Mary Jean Amon, Rakibul Hasan, Kurt Hugenberg, Bennett I Bertenthal, and Apu Kapadia. Influencing Photo Sharing Decisions on Social Media: A Case of Paradoxical Findings. In the Proceedings of the IEEE Symposium on Security & Privacy (**Oakland '20**).

[6] Rakibul Hasan, Bennett I Bertenthal, Kurt Hugenberg, and Apu Kapadia. Your Photo is so Funny that I don't Mind Violating Your Privacy by Sharing it: Effects of Individual Humor Styles on Online Photo-sharing Behaviors. To appear in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (**CHI'21**).

[7] Rakibul Hasan, Cristobal Cheyre, Yong-Yeol Ahn, Roberto Hoyle, and Apu Kapadia. The Impact of Viral Posts on Visibility and Behavior: A Longitudinal Study of Scientists on Twitter. In review.