

SOCIAL MEDIA PLATFORMS, THE RISKS OF MASS ATROCITIES, AND OPPORTUNITIES FOR ATROCITY PREVENTION

2023 Sudikoff Interdisciplinary Seminar on Genocide Prevention



UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE

THE UNITED STATES HOLOCAUST MEMORIAL MUSEUM

teaches that the Holocaust was preventable and that by heeding warning signs and taking early action, individuals and governments can save lives. With this knowledge, the **Simon-Skjodt Center for the Prevention of Genocide** works to do for the victims of genocide today what the world failed to do for the Jews of Europe in the 1930s and 1940s. The mandate of the Simon-Skjodt Center is to alert the United States' national conscience, influence policy makers, and stimulate worldwide action to prevent and work to halt acts of genocide or related crimes against humanity, and advance justice and accountability. Learn more at ushmm.org/genocide-prevention.

TALLAN DONINE, Research Assistant at the Simon-Skjodt Center for the Prevention of Genocide.

COVER: 27.01.2017, Yangon, Republic of the Union of Myanmar, Asia - Workers sit in front of a Buddhist temple during their break and use their smartphones. *Olaf Schuelke / Alamy Stock Photo*

CONTENTS

Introduction.....	1
The Relationship Between Social Media and Mass Atrocities.....	2
Social Media Platforms and the Prevention of Mass Atrocities.....	4
Endnotes.....	10

INTRODUCTION

On January 19 and 24, the Simon-Skjodt Center for the Prevention of Genocide at the United States Holocaust Memorial Museum brought together a diverse group of scholars, representatives of social media companies, practitioners, and policy makers for two 90-minute virtual convenings to discuss social media platforms, the risks of mass atrocities, and opportunities for atrocity prevention. Motivations for this seminar included (1) taking stock of knowledge about the relationship between social media, mass atrocities, and atrocity prevention; (2) generating new ideas for future research or practice on this topic; and (3) exploring how the Simon-Skjodt Center can play a constructive role in this area.

This rapporteur's report summarizes key observations raised during both discussions under the Chatham House rule of non-attribution and does not necessarily represent the views of all participants.

The first session focused on exploring the interaction between social media platforms and risks of mass atrocities. Guiding questions included:

- How might social media platforms contribute to risk factors associated with mass atrocities?
- How might perpetrators use social media platforms to enact mass atrocities?
- How might the relationship between social media platforms and mass atrocities change in the future?

The second session focused on social media platforms and the prevention of mass atrocities. Guiding questions included:

- How do technology firms and others understand the risks that social media platforms may inadvertently contribute to the commission of mass atrocities?
- What steps have firms taken to date to address these risks?
- What new or different actions can social media companies take to design and manage their platforms in ways that would help reduce risks of mass atrocities?
- What type of regulation of social media platforms would most effectively balance interests in civil liberties and violence prevention?

THE RELATIONSHIP BETWEEN SOCIAL MEDIA AND MASS ATROCITIES

In recent years, a disturbing pattern across cases such as Sri Lanka, Burma, India, and Ethiopia has raised alarm surrounding social media's influence on large-scale, group-targeted violence. However, research about the specific relationship between social media platforms and the onset and escalation of mass atrocities remains relatively limited.¹

Throughout both sessions, several participants voiced the tension between the need for more research and the need for action—reflecting a recurring tension in the broader atrocity prevention field. Some participants expressed concern that a limited understanding of causal factors might unintentionally lead to counterproductive solutions. Additionally, multiple participants indicated that identifying more effective ways to respond further upstream requires understanding the causal mechanisms at play. Other participants argued that social media's effects on atrocity risks are plain and that striving to meet academic standards of knowledge could distract from the urgency of acting to address risks. Participants generally agreed that while researchers need more data to understand better the relationship between social media and mass atrocities, this should not prevent responses to present social media risks. Instead, research and policy responses could simultaneously reinforce one another.

Research challenges

Participants generally agreed that social media's effects on mass atrocity risks are likely to vary across high-risk contexts. As one participant outlined, the differences in platform functions and their effects are becoming more salient as social media environments (within countries and between social groups) become more fragmented. These different contexts and evolving environments increase the need to develop more extensive qualitative data on social media and user behavior through specific case studies. Yet, datasets necessary for researchers to conduct adequate causal work connecting online harms to offline consequences remain unavailable or inaccessible.

Multiple participants called on social media companies to transparently and publicly share data with researchers in a way that protects user privacy and security. For example, one participant expressed concern that social media companies have not released relevant data about social media use during the Rohingya genocide more than five years after the escalation of large-scale anti-Rohingya violence. This has limited researchers' ability to examine the nature and extent of social media's influence on the mass atrocities in this case and distill lessons to carry forward in other comparable contexts. Additionally, multiple participants noted that while some social media companies have conducted causal research on related questions internally, these companies have not released their full findings or data to enable replication.

Participants agreed that the multiple pathways by which social media may influence mass atrocity risks complicate this research. For example, one participant observed that the interaction between social media

and other types of media makes it difficult to isolate social media's effects on mass atrocity outcomes. One participant also cited research that indicates that social media content can have a more substantial effect when it echoes existing conversations offline ([Lupu et al. 2023](#)).

Participants also addressed temporal challenges. Since social media appears to affect users gradually over time, it is harder to capture its immediate and long-term effects on mass atrocity risks. Additionally, the rapid speed at which content can spread on social media poses challenges for measuring its effects. Multiple participants noted it is difficult to measure how many people are exposed to any particular content. That is, outside of interactions recorded by retweets and likes, many more users view content and may be influenced by it than these metrics reflect.

Ideas for future research

Multiple participants called for expanded **interdisciplinary research and collaborative partnerships between researchers, social media companies, and civil society advocates**. Considering social media and its effects on mass atrocity risks are likely to vary across high-risk contexts, one participant suggested that researchers work towards consensus about the potential mechanisms by which social media influences mass atrocity risks. Participants also suggested that research should link these mechanisms to operational examples. Participants added that researchers could better assess causality by examining the effects of social media in different social and political contexts. Further, participants said that we need more research on the effects of social media in contexts other than the United States.

One participant proposed three potential mechanisms through which social media may influence mass atrocities: coordination (through information sharing), activation (of people who already hold dangerous views), and persuasion (changing people's underlying preferences). Another participant suggested viewing social media as an intermediary factor that may accelerate action for good or bad. For example, while social media may accelerate the spread of harmful content, it may also assist in protecting at-risk groups, such as through end-to-end encrypted messaging.

Several participants suggested research should focus on **which actors might use social media platforms to help commit mass atrocities or accelerate the risks of mass atrocities**. For example, one participant recommended identifying and examining the incentives of elites or "political entrepreneurs" who may have greater capability to accelerate the spread of harmful content. Research on this may assist with early warning for mass atrocities by focusing on potential perpetrators and their incentives.

Additionally, one participant noted the potential benefits of **differentiating social media users** more extensively. They pointed out that the general public represents a significantly different audience for social media content than government or military personnel. Therefore, the ways social media users might encourage violence or contribute to risks will likely play out differently in these groups.

Multiple participants **suggested research should also address risks posed by social media beyond hate speech**. Specifically, research could focus on less extreme forms of dangerous speech that could cause

offline harm but are not easy to detect because they may be more subtle. In addition, participants suggested research should address other, more difficult-to-quantify product-based risks. For example, perpetrators might use community-oriented features such as ‘Friends’ lists or group pages to identify and target people. To this effect, one participant noted Facebook created a feature to [hide](#) users’ “Friends” lists in Afghanistan amid the August 2021 Taliban takeover in an attempt to limit targeting of vulnerable groups.

Participants also mentioned that different evidentiary standards and research are necessary for different kinds of action: political, legal, and criminal. One participant suggested that a tension exists between the desire to act preventively and the fact that evidence about the role of social media is typically clear only after atrocities have been committed. It could be useful, therefore, to consider in advance of a crisis the tools available to address potential harms posed by social media and the research or evidence that would be deemed sufficient to justify the use of different tools.

Participants proposed additional research questions:

- How is the relationship between social media and mass atrocity risks affected by different regime types and capacities?
- How are different communities or types of groups targeted through social media? Are some groups more vulnerable than others?
- What is the relationship between a country’s per capita social media penetration and its level of mass atrocity risks?
- What civil society resources and resiliencies might address risks posed by social media in at-risk environments?
- How do mass atrocity risks compare across private and public social media platforms? How does information flow across platforms for different purposes?

Finally, one participant said it is important for researchers to recognize and account for how their social position, identities, and power can affect research.

SOCIAL MEDIA PLATFORMS AND THE PREVENTION OF MASS ATROCITIES

Participants voiced differing opinions about how central business incentives are for determining how seriously social media companies work to help prevent atrocities.

One participant challenged the common conception that social media companies seek to propel content and maximize revenue regardless of its potential for harm. The participant commented that social media companies’ business models do not benefit from harmful content. Rather, they noted dangerous content can harm business interests by deterring advertisers. Multiple participants noted social media companies

generally do not want their products to contribute to online or offline harm. However, a participant explained that while it may not be in the company's interest to promote potentially harmful content, interest alone will not prevent adverse outcomes.

One participant proposed that social media companies may lack incentive to focus on mass atrocity risks because market presence may not be particularly high in places with significant mass atrocity risks.

However, some social media companies seek to identify countries at high risk of mass atrocities to help steer resource allocation. For example, a participant described how one major social media company has identified at-risk countries and subsequently increased local language content moderation, worked with third-party content moderators, and advanced efforts to educate users on digital security in some cases.

Potential opportunities for atrocity prevention

Several participants expressed frustration that the attention on social media often occurs after the onset of violence. Multiple participants agreed social media companies should invest more in identifying and assessing atrocity risks and acting preventively.

Building healthy social media ecosystems

Because people generally seek to use social media constructively, several participants agreed that social media companies should seek to build healthy social media “ecosystems.” Participants indicated that a healthy social media ecosystem is an environment that supports communities by fostering constructive and safe interaction for all users.

Participants suggested it is important to explore how to encourage social media companies to develop these ecosystems. Multiple participants agreed safe and transparent data access is necessary for researchers to examine what it would take to create healthy ecosystems. However, one participant voiced concern that online environments reflect offline environments, citing research that indicates social media reflects offline polarization ([Barberá 2020](#)). Additionally, the participant noted building these healthy ecosystems may be more difficult if certain actors continue to find nefarious ways to use the platforms. One participant suggested capacity building and training in the non-harmful use of social media could contribute to developing these healthy ecosystems.

Participants also suggested further exploring the potential for social media users to promote content that might de-escalate conflicts. Another participant added that while the conversations often focus on influential actors using social media for nefarious means, influential actors may also promote more positive content, or what is referred to as “counterspeech.” The participant suggested that such an effort to encourage positive content would require extensive consultations with diverse civil society groups to assess which key stakeholders may help mitigate potentially dangerous situations. However, another participant cautioned that efforts to shape online discourse can become highly political and controversial. One participant suggested it is important to test different content-promoting approaches in actual cases to gauge their effectiveness.

Supporting the role of local actors

Multiple participants agreed it is important, although difficult, to bridge the gap between social media companies and communities on the ground in at-risk environments, especially once a crisis begins.

One participant said that local civil society groups need to be included in research on new and existing prevention approaches, particularly to help define what it means for these approaches to be effective. The participant explained that researchers could collaborate with community groups to identify measurable markers of a healthy ecosystem to test out these practices and accumulate research. An additional participant noted other forms of political crisis and conflicts, such as mass protests, may serve as a testing environment for this work. While social media companies may be conducting similar research already, one participant suggested it is neither rigorous nor publicly accessible.

Participants discussed challenges associated with social media platforms' reliance on trusted local partners to raise potential issues. Social media platforms may over-rely on these communities and thus over-burden them, raising concerns about the need to protect these groups and expand partnership networks. Further, one participant suggested stakeholders should seek a deeper understanding of who should participate in conversations focused on identifying mass atrocity risks. A participant suggested all parties also need to interrogate who makes these decisions.

One participant said solutions should not focus only on what social media companies or governments can or should do to mitigate risks related to social media platforms. Instead, they suggested solutions should also draw on ideas from other parties, including grass-roots, user-led efforts to counter harm.

Another participant said that local actors constantly seek ways to address the mass atrocity risks associated with social media platforms. The participant suggested groups could create and share a playbook documenting best practices, response strategies, associated risks, and examples for local actors to address harm linked to social media at different stages. In a similar vein, the Simon-Skjoldt Center published [*Defusing Hate: A Strategic Guide to Counteract Dangerous Speech*](#) (2016) to provide activists, religious and civil society leaders, and their supporters with strategies and tools for limiting the influence of dangerous speech. This guide may provide a framework for similar work focused on harm related to social media.

Advancing due diligence

Several participants suggested that social media companies should undertake comprehensive due diligence before entering markets in at-risk environments. Multiple participants agreed it is important for social media companies to account for how people may use their products in areas where the companies do not have a market presence, but where a large part of the population uses their services. Additionally, one participant suggested social media companies' "duty of care" responsibility extends to people who directly use the product and those who may be affected by the actions of product users. This requires, as

one participant indicated, that social media companies consider the human rights effects if everyone everywhere has access to their products.

Existing international frameworks serve as potential starting points for social media companies to engage in more robust human rights due diligence, prevention, and response to mass atrocity onsets and escalation. One participant noted that the [United Nations Guiding Principles on Business and Human Rights](#) provide a core set of human rights standards for social media companies to incorporate into their risk mitigation work. Additionally, multiple participants cited the [United Nations Rabat Plan of Action](#) as providing a useful threshold test for determining whether speech “constitutes incitement to discrimination, hostility or violence” and should therefore be prohibited by law. The Rabat Plan encourages consideration of several elements, among them the “status of the speaker,” the surrounding “social and political context,” and the content’s “likelihood of harm, including imminence.”² Participants noted the Rabat Plan can advance prevention strategies by accounting for the context and potential cumulative impact of harmful speech.

One participant suggested that social media platforms should be required by law to meet a “minimal technical standard” before entering markets in “mass atrocity capable environments.” According to the participant, this could help ensure social media companies have taken mass atrocity risks into account and emphasizes the need for proactive responses to risks or onsets should they arise.

Strengthening responsible product design

Participants suggested social media platforms—specifically engineers and data scientists—should address potential mass atrocity risks associated with product design, such as group features like ‘Friends’ lists previously discussed. Before social media companies launch products, they could incorporate local perspectives into atrocity prevention-oriented designs and testing, for example. A participant recommended social media companies should maintain ongoing engagement with local actors to monitor the evolution of features and potential unanticipated issues stemming from the products.

One participant asked what it would take to inculcate anti-mass atrocities value into social media companies' product design processes. The participant cited the [IEEE Standard Model Process for Addressing Ethical Concerns During System Design](#) as a way to embed values into technical systems at the outset, which can contribute to significant change over time and harm prevention.

The participant also raised additional questions about how to design social media platforms that advance “anti-atrocity” values—such as restraint and cooperation—amid calls for violence. To guide the development of products in line with these values, the participant encouraged social media companies and civil society organizations to develop metrics and goals for improving the resilience of social media platforms that go beyond individual engagement with anti-atrocity content.

Addressing legal responsibility and liability

One participant said that social media companies should share relevant data with United Nations investigative mechanisms and international courts, such as in the case of Burma. They also suggested social media companies should work to preserve evidence relevant to mass atrocity cases to ensure it can be used in legal settings where applicable.

Additionally, multiple participants agreed it is difficult to establish criminal liability for social media companies that allow their platforms to be used to spread hate speech, incite violence, or coordinate violent attacks in cases such as Burma. One participant also said more research is needed on causal mechanisms to support these legal processes.

Exploring government regulation

As the seminar background paper outlines, in the United States, calls for addressing social media risks have centered on reforming [Section 230](#) of the 1996 Communications Decency Act. The provision holds that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." One participant noted that some social media companies view Section 230 as providing critical protection that enables content moderation and that reducing these protections could limit their ability to remove potentially harmful content. A participant countered that while Section 230 provides social media companies important legal protections, it has negatively affected the development of a "duty of care" standard. Multiple participants suggested that changes in government regulation, such as adding criminal and legal incentives, are necessary to promote a "duty of care" standard among social media companies.

One participant suggested some social media companies support evolution in regulatory frameworks, such as the European Union's [Digital Services Act](#) provisions to increase transparency and share data with authorities and researchers.

Accounting for potential unintended consequences

One participant suggested that if atrocities unfold with perceived links to social media activity, then policy makers and social media platforms should consider taking tactical responses geared at containment, such as shutting down site access. However, as multiple participants responded, research has linked internet blackouts to increased repression ([Gohdes 2015](#)). These responses highlighted the importance of assessing the potentially dangerous unintended consequences that may come from prevention approaches before enacting those responses.

Enhancing collaboration between stakeholders

Participants suggested better collaboration between stakeholders writ large: specifically, better collaboration between social media companies, researchers, and civil society groups to advance research on the relationship between mass atrocities and social media and research into potential policy, programmatic, and design responses to mass atrocity risks. These groups could strengthen efforts to

prevent potential social media-related harm. For example, one participant suggested social media companies would appreciate feedback from civil society and researchers on how to create spaces for users to exercise their rights to free speech and reduce the harm of negative content across platforms. Multiple participants agreed that collaboration between social media companies can enable better risk monitoring across platforms.

One participant suggested that adopting a risk-assessment model like [Brightline Watch](#) could be useful to assess opinions on potential social media prevention approaches and trends among researchers, civil society practitioners, and other groups more regularly.

ENDNOTES

¹ The Seminar’s background paper surveys current knowledge and identifies important gaps in understanding about (1) how social media platforms have contributed to the risk and occurrence of mass atrocities in the past and how they might do so in the future; and (2) strategies to help prevent social media from fueling mass atrocities. For more information, see

https://www.ushmm.org/m/pdfs/2023_Sudikoff_Interdisciplinary_Seminar_on_Genocide_Prevention_-_Background_Paper.pdf.

² “One-pager on “incitement to hatred,”” Office of the United Nations High Commissioner for Human Rights, accessed January 2023, https://www.ohchr.org/sites/default/files/Rabat_threshold_test.pdf.

A living memorial to the Holocaust, the **UNITED STATES HOLOCAUST MEMORIAL MUSEUM** inspires citizens and leaders worldwide to confront hatred, prevent genocide, and promote human dignity. Its far-reaching educational programs and global impact are made possible by generous donors.



ushmm.org/connect

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

**SIMON-SKJODT CENTER
FOR THE PREVENTION OF GENOCIDE**

100 Raoul Wallenberg Place, SW Washington, DC 20024-2126 ushmm.org