



Globally Scaling Digital Solutions for Managing Misinformation.

The Sentinel Project



THE SENTINEL PROJECT





Contents

Scaling Digital Solutions for Managing Disinformation.	1
Acknowledgements	3
Executive summary	4
Section 1: Introduction	5
Aims and scope.	5
Methodology	5
Main findings and discussion	6
Section 2: SP's Misinformation Management Methodology and Research Methods	9
SP's misinformation management framework.	9
SP's approach to research	12
Glossary of abbreviations and definitions	14
Section 3: Main Findings – Part 1	18
Identification of misinformation.	18
The distinction between online and offline misinformation	18
The creation, speed, breadth, and design of misinformation spread.	19
The variety of sources of information	20
Misinformation is profitable	21
The effects of misinformation are not well understood	22
How populations are targeted for misinformation campaigns	22
Demographic factors	22
National politics and elections	23
The effects of misinformation on democracy.	25
The relationship between misinformation, hate speech, and atrocities	26
Violence and viruses.	29
Misinformation verification and establishing facts – What do we do to counter misinformation?	30
Examples of joint content flagging initiatives	31
Examples of universities, public institutions, news media, social media companies	31
Flagging COVID-19 misinformation	32
Content verification, computational approaches, and AI-based initiatives	34
Case Study 1 - Topic analysis of rumours database across all projects since 2014	36
Case Study 2 - Una Hakika in Kenya.	37
Case Study 3 - Hagiga Wahid and Uganda and South Sudan.	38
Case Study 4 - Kijiji Cha Amani in the Democratic Republic of the Congo	39
Section 3: Main Findings – Part 2	42
Technology companies	42
Government interventions	43
Elections	44
Media literacy	44
Civil society	45
Humanitarian organizations	46
Crowdsourcing	46
Section 5: Recommendations	48
Works cited	51



Acknowledgements

The Sentinel Project is a non-profit organization dedicated to assisting communities worldwide that are threatened by mass atrocities—through direct cooperation with the people in harm's way and the innovative use of technology. The Sentinel Project currently has active field operations in Kenya, Uganda, South Sudan, and the Democratic Republic of the Congo, all of which focus on countering the spread of harmful misinformation that contributes to conflict and violence.

This report was made possible by funding from the International Development Research Centre (IDRC) and through the contribution of expertise from people working in the atrocity prevention, peacebuilding, humanitarian aid, international development, and technology fields. In addition to thanking IDRC for its financial support, the Sentinel Project wishes to thank all of those who provided invaluable insights which contributed to this report.





Executive summary

This report examines the related phenomena of rumours and misinformation—which includes disinformation—especially in relation to how they contribute to conflict, violence, and societal instability. The team that compiled this report addressed this complex and multifaceted issue by focusing on answering the following two main research questions.

1. How can misinformation management effectively and sustainably operate at scales involving mass data quantities and audiences which cannot be engaged on an interpersonal level?
2. How can social media platforms and related government policies reduce the ease of proliferation for misinformation and the degree to which it has negative impacts?

In producing this report, the research team has offered a snapshot of misinformation management efforts worldwide by compiling and synthesizing findings from research articles, online reports, and practical initiatives that address this topic. The report uses case studies from various initiatives operated by the Sentinel Project (SP) to highlight how misinformation management frameworks can be replicated and scaled in different contexts. Additionally, the research team examined solutions that have been implemented by various other organizations, technology companies, and governments. The report ends with practical recommendations for technology companies and national governments on how to approach misinformation management.

While the research team found that it is difficult to create any single approach which is universally applicable for addressing the problem of misinformation, there are certainly several themes and principles that can guide most efforts. Following are the themes, principles, and general conclusions of this report.

1. Rumours and misinformation are fundamentally human phenomena, so any approach to

countering them must take human factors into account to be effective.

2. Technological tools can be very useful both for those who seek to spread misinformation and for those who seek to counter it.
3. Technology and human moderators must complement each other in misinformation management efforts since there is no single technological solution to this human problem.
4. Governments have an important role to play in addressing misinformation but this must be done with restraint and in balance for fundamental rights and freedoms.
5. Civil society actors are critical for effectively addressing misinformation, since they may have the best understanding of the nuanced contextual factors that affect the relevant populations.
6. Technology companies must assume more responsibility for monitoring and moderating misinformation on their platforms.
7. Further research is required to understand the impact of rumours and misinformation, as well as their relationships with hate speech and physical violence.



Section 1: Introduction

Recent years have seen a dramatic and alarming increase in the prevalence of misinformation worldwide, especially as it proliferates online through various social media channels.

As the 2019 Atlantic Council report *Disinformation in Democracies: Strengthening Digital Resilience in Latin America* states,

Governments, technology and social media companies, civil society, and media don't yet fully understand the challenges disinformation poses to democracy. These stakeholders, each with different sets of motivations, do not yet fully see eye to eye on the best way to address the rise of influence operations online. Meanwhile, regulatory frameworks remain outdated as technologies evolve faster than the laws that regulate online abuses and punish bad actors.

While the phenomenon of misinformation is not new, the degree of damage that it can cause in an increasingly digitized and interconnected world is unprecedented. The new elevated threat of misinformation is due to a combination of factors, including the speed of creation and dissemination, the variety of sources, its profitability, and people's levels of vulnerability (BBC, 2018; The Guardian, 2016; Lazer et al., 2018; Wheeler, 2017; Shu et al., 2020). As noted below, this threat is likely to continue growing as the cost of generating such content and disseminating it to very large and sometimes international audiences decreases. Artificial intelligence (AI) and machine-generated misinformation can also increase the threat level from malicious actors who create and spread disinformation (DiResta 2020; Kertysova, 2018). Furthermore, as shown through the examples and case studies provided in this report, while most observers have recently focused their attention on prominent misinformation cases threatening established democracies, it is also destabilizing countries

throughout the Global South. Political processes, elections, and international cooperation have all been affected by misinformation. Importantly for SP's work, there is also a relationship between misinformation, hate speech, and mass atrocities. The proliferation of misinformation has only been further highlighted by public health crises. For example, the ongoing COVID-19 pandemic has been exacerbated worldwide by rumours and misinformation that hinder public health responses, and by misinformation produced by political actors seeking to exploit the fear and uncertainty that some populations are facing. Consequently, understanding and addressing misinformation is a global problem that requires novel approaches—mainly due to the extent of social media magnification in recent years.

Aims and scope

Misinformation is a global phenomenon that needs to be better understood in order to diminish its pernicious and potentially violent effects. This paper investigates the breadth of the problem of misinformation on a global scale—especially as it relates to hate speech and violence—and looks at some current initiatives that attempt to address this problem.

Methodology

Two sets of operational and research methods guide this paper and inform each other. Regarding operating procedures, SP's on-the-ground misinformation management framework directs our efforts to prevent and mitigate atrocities in various countries. As will be detailed in Section 2, the overall misinformation



management process consists of three basic stages: monitoring, verification, and counter-messaging.

- Monitoring focuses on identifying rumours and misinformation circulating amongst a given population, typically achieved through a crowdsourcing approach that gathers rumour reports from the general public through channels like SMS (text messaging), voice calls, social media, and volunteer proxies.
- Verification involves investigating rumours with the help of a network of trusted stakeholders, consulting multiple sources, and exercising judgement.
- Counter-messaging focuses on returning verified information and other relevant details to communities using the same monitoring channels through which rumour reports are submitted.

Regarding this report, a standard research methodology was used to understand and analyze the extent of misinformation and evaluate the tools and policies used to mitigate misinformation. The first phase of this research involved a literature review of recent news articles, reports, and academic resources on misinformation management. The second phase of the study involved developing specific case studies of SP's misinformation management projects. In this phase, the research team brought together previously gathered data and examined established project protocols and materials that have been used in SP's projects. Together, these findings provide a solid foundation from which to build technological and governmental policy recommendations.

Main findings and discussion

The literature review and case study analysis revealed five salient points.

1. The literature review and case study analysis revealed five salient points.
2. Online misinformation has characteristics

that make it distinct from other forms of misinformation. The creation, speed, breadth, design, and profitability of online misinformation make it a phenomenon of much greater danger than its offline counterparts.

3. Although research in this area is still nascent and requires significant further work to draw definitive conclusions, it is clear that some people are more susceptible to misinformation and that some populations are specifically targeted for disinformation campaigns. These two factors mean that it is important to understand the factors that make a given population vulnerable to misinformation.
4. Misinformation disrupts societies by creating tensions in social relations and may be linked to hate speech and physical violence. It also disrupts democracies by undermining institutional trust, interfering with elections, and exacerbating confusion around various issues.
5. Initiatives to counter misinformation must span all facets of society and be joint initiatives between governments, public institutions, and private sector companies that flag content and home in on misinformation sources. These recent efforts to operationalize macro-level misinformation management need to account for a considerable number of factors. This reality highlights the need for misinformation management systems that can effectively and sustainably operate at scales involving mass data quantities. However, these top-down, highly technical initiatives are not a panacea. Low-technology grassroots initiatives remain important due to misinformation management's contextual nature—which requires intensive trust building—especially in contexts that are already unstable and conflict ridden.
6. The case studies outlined here demonstrate that local-level initiatives can be scaled up to reach larger populations and can also be replicated in different contexts while incorporating important contextual factors. Furthermore, both funders



and implementers need to recognize that these initiatives can be resource-intensive efforts so having a significant impact requires substantial investments.

We highlighted three central considerations in response to the first question that guided this research, “How can misinformation management effectively and sustainably operate at scales involving mass data quantities and audiences which cannot be engaged on an interpersonal level?”

1. Human input into the misinformation management process is critically important and likely will continue to be for the foreseeable future since such efforts are labour intensive and many of their aspects cannot be automated.
2. Misinformation management systems cannot only be imposed from above. Instead, they must be implemented by entering into communities using culturally relevant processes followed by cooperative efforts.
3. Trust is an essential element of any misinformation management project. Thus, a project will immediately lose most of its value to the community and the pursuit of peace if trust is not built. Suspicion or hostility from some community members is to be expected. Misinformation management often involves controversial and emotive issues, and not everyone in a community will immediately appreciate these initiatives. Indeed, SP's projects deal with intricate social dynamics in local contexts and highlight the importance of user trust. Any attempts at scaling up or reconciling micro and macro approaches should consider these points since such considerations make it challenging to universalize projects that require intensive on-the-ground knowledge of the local context.

There are also three main considerations in response to the second question, “How can social media platforms and related government policies reduce the ease of

proliferation for misinformation and the degree to which it has negative impacts?”

1. Corporations that operate social media platforms often have more legal freedom than governments to engage in surveillance and censorship. However, even when inclined to address misinformation, such companies struggle to find an appropriate global standard since their users typically span many (potentially hundreds) of legal and regulatory jurisdictions as well as cultural contexts.
2. Governmental policy interventions remain central to reducing misinformation proliferation, which increasingly propagates on social media platforms controlled by the private sector. However, there is no one-size-fits-all answer since there is a high level of variation between jurisdictions in terms of political conditions, regulations, and norms.
3. Managing misinformation through policy or law introduces a tension between the desire to reduce the prevalence of the dangerous phenomenon of misinformation while also preserving the right to freedom of expression. Once discussions about countering misinformation include government intervention, there are genuine questions that need to be asked about whether governments should play a central role in developing regulation around misinformation, due to the potentially deleterious impacts on freedom of expression and related rights.





Section 2: SP's Misinformation Management Methodology and Research Methods

The motto “peace begins with the truth” guides all of SP’s misinformation management projects and it is derived from the recognition that harmful rumours and misinformation can drive the distrust, fear, and hatred that enable violent conflict and atrocities. So, how does one ascertain the truth in challenging environments like the Democratic Republic of the Congo, Uganda,

Kenya, and South Sudan? This section on methodology and research methods contains two main components. First is an examination of SP’s hands-on approach to misinformation management. Second is the research team’s approach to data gathering and analysis for this report.

SP’s misinformation management framework

First, we will elaborate upon the misinformation management framework that guides SP’s efforts to prevent and mitigate violence and mass atrocities. As introduced briefly above, this framework is incorporated into various projects and generally follows three stages which are adjusted to fit specific contexts: monitoring, verification, and counter-messaging.

Monitoring

Monitoring typically focuses on crowdsourcing to gather rumour reports from the general public through channels like SMS (text messages), voice calls, social media, and volunteer proxies. It is also possible to incorporate automated monitoring of data sources such as social media if this is relevant and technologically feasible in a given context.





Verification

Once a rumour report is received, it is added to a database called WikiRumours, which enables geographically distributed teams to collaborate on the rumour verification process. Rumours are prioritized and assigned to human moderators for investigation, which has three main elements, as outlined below.

- 1. Consult a network of trusted stakeholders** – It is critical in every misinformation management deployment to establish a network of trusted information sources by building information-sharing relationships with key stakeholders. The project team completes an information verification matrix (IVM) to guide this process, which organizes sources by subject matter expertise and geographical coverage. This exercise is repeated periodically to update the IVM and maintain it as a living document. The project team engages a wide variety of civil society sources, local media outlets, international NGOs, United Nations agencies, government officials, local leaders, volunteers, and others. The IVM identifies gaps in the trusted stakeholders network and serves as a reference point for team members to quickly consult information sources (including both individuals and organizations) that can provide information related to a given rumour.
- 2. Consulting multiple sources** – The project team consults multiple sources when drawing on the trusted stakeholders network to mitigate bias, gain various perspectives, and gather as much information as possible. It is important to note that bias is usually not conscious and sources are not suspected of intentionally providing misleading information. Instead, it is acknowledged that different actors may have access to distinct types of data and have various perspectives on a given rumour or incident. For example, a community leader, humanitarian worker, and government security official are all likely to know different things about a case

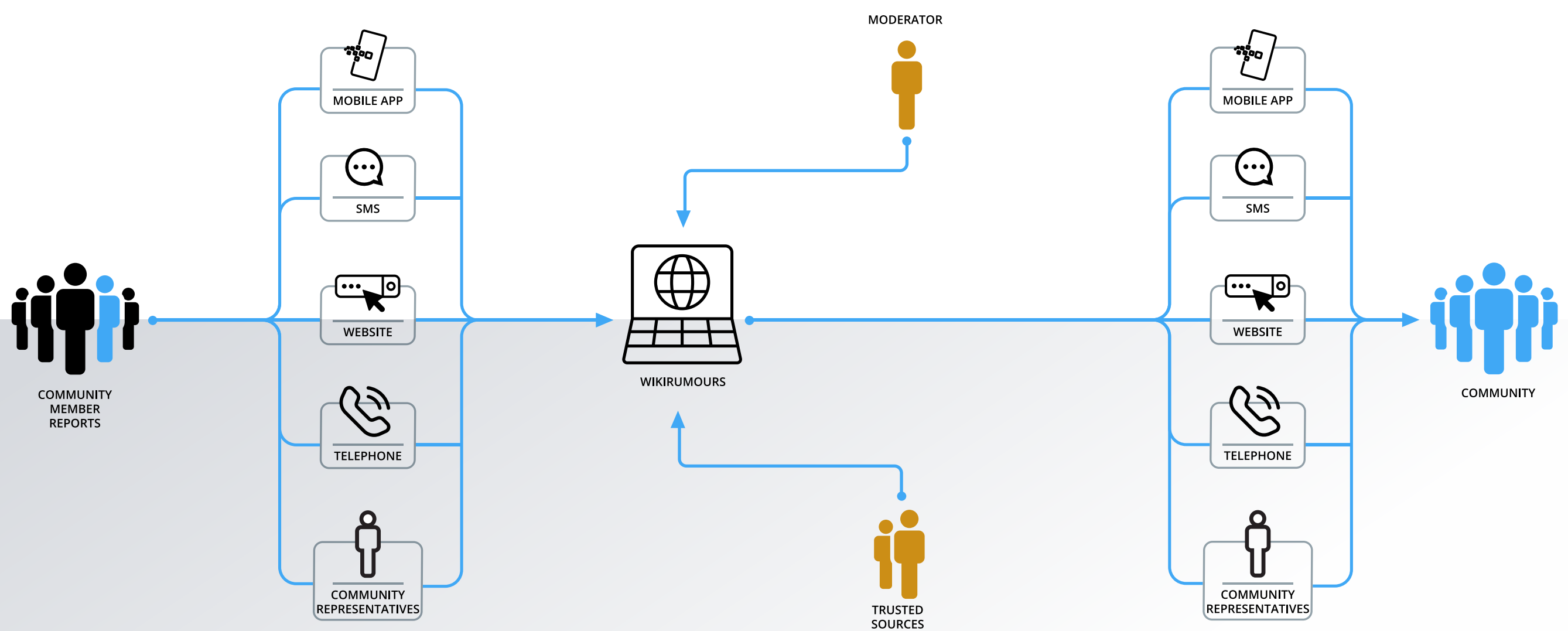
of intercommunal conflict and have different perspectives on it. Recognizing this difference from the outset improves the project team's analysis as it seeks to establish the facts.

- 3. Exercising judgement** – Of course, the information gathered from even the most trusted sources may not always enable a clear-cut determination of whether a given rumour is true or false. In the case of conflicting evidence, the project team may need to exercise a degree of judgement and decide where a given rumour falls between the "confirmed true" and "confirmed false" ends of the verification status spectrum. For this reason, the SP model provides room for assigning other statuses such as "probably true" and "probably false." In such cases, project team members consider factors such as plausibility and precedent when making determinations. Ambiguous cases also require engaging stakeholders in the counter-messaging stage that follows verification since they can better deliver more nuanced responses to beneficiaries.

Counter-messaging

Counter-messaging focuses on returning verified information and other relevant details to beneficiary communities using the same monitoring channels through which a given project receives rumour reports. SP further uses counter-messaging for issuing localized warnings and public alerts during ongoing crises to help people navigate dangerous situations.

Relevant case studies have been placed throughout this report to demonstrate how the rumour verification process works in practice. The WikiRumours database is detailed in Case Study 1, which shows how this purpose-built software helps to organize the rumours that are reported through each of SP's currently ongoing projects (Una Hakika, Hagiga Wahid, and Kijiji Cha Amani).





SP's approach to research

This section outlines the methodology and methods used to develop the findings of this report. As mentioned in the introduction, SP has focused on misinformation management as a critical topic of concern for both impact-oriented projects and academic research. To understand this increasingly important topic, the project team applied a standardized research methodology to identify and analyze the phenomenon.

The goal of this research was to investigate the breadth of misinformation and map out the current initiatives being undertaken to mitigate it. The research questions were:

1. How can misinformation management effectively and sustainably operate at scales involving mass data quantities and audiences which cannot be engaged on an interpersonal level?
2. How can social media platforms and related government policies reduce the ease of proliferation for misinformation and the degree to which it has negative impacts?

Answering these questions required information from multiple sources, which was gathered and analyzed in two phases. The first phase involved a literature review of recent news articles, reports, and academic resources on misinformation efforts to counter it. We included literature from 2016 to 2020 with reasonably broad inclusion criteria. If an article discussed misinformation in any form and across any geographic area then it was included. The SP team created a spreadsheet of all sources, which were then sorted thematically by closely examining the information in the resources to identify broad themes. This analysis determined that the articles fell into at least three groups:

1. Identification of misinformation/gathering data about misinformation
2. Establishing facts
3. Countering misinformation

These three themes reveal the chronological development of misinformation and its effects, and we use this categorization to structure our findings below.

The second phase of this research involved developing specific case studies of misinformation management projects. In this phase, we brought together previously gathered data such as project protocols and materials from SP's on the ground projects. Regarding the analysis, this step included studying SP's projects from the lens of the three themes identified above. Brought together, these two phases offer a holistic lens into the prevalence of misinformation, and the current state of misinformation management, particularly in the Global South.





Glossary of abbreviations and definitions

AI	Artificial intelligence	UN	United Nations
EU	European Union	UNGA	United Nations General Assembly
IFCN	International Fact-Checking Network	UNAIDS	Joint United Nations Programme on HIV and AIDS
IIFMM	Independent International Fact-Finding Mission on Myanmar	UNICEF	United Nations International Children's Emergency Fund
MIT	Massachusetts Institute of Technology	UNDP	United Nations Development Programme
NATO	North Atlantic Treaty Organization	UNESCO	United Nations Educational, Scientific and Cultural Organization
NGO	Non-governmental organization	WHO	World Health Organization

Table 1

Cyber-troops or troll armies	The Oxford Internet Institute defines cyber troops as “government or political party actors tasked with manipulating public opinion online” (Bradshaw & Howard, 2017 in Bradshaw & Howard, 2018, p. 4).
Digital authoritarianism	<p>Brookings defines digital authoritarianism as “the use of digital information technology by authoritarian regimes to surveil, repress, and manipulate domestic and foreign population” (p. 1). Erol Yayboke and Sam Brannen of The Center for Strategic and International Studies define digital authoritarianism in their 2020 recent brief as:</p> <p>...the use of the Internet and related digital technologies by leaders with authoritarian tendencies to decrease trust in public institutions, increase social and political control, and/or undermine civil liberties. Human rights and civil liberties are at risk, including freedom of movement, the right to speak freely and express political dissent, and the right to personal privacy, online and off. Digital authoritarianism co-opts and corrupts the foundational principles of democratic and open societies; its goal is not just to break them down, but to redefine and reshape them in their authoritarian image. (p. 2)</p>

Disinformation	<p>The independent high-level group on fake news and online disinformation of the European Commission (2018) defines disinformation “as false, inaccurate, or misleading information designed, presented, and promoted to intentionally cause public harm or for profit” (p. 10). A Council of Europe Report by Wardle and Hossein (2017) refers to “information disorder” as the best terminology to capture the phenomenon that includes misinformation, disinformation, and malinformation. According to the authors (p. 5), these are defined as follows:</p> <ul style="list-style-type: none"> • Misinformation is when false information is shared but no harm is meant. • Disinformation is when false information is intentionally shared to cause harm. • Malinformation is when truthful information is shared to harm. <p>In consideration of Wardle and Hossein’s (2017) report, SP’s report uses the term disinformation to refer to any false information that is deliberately spread. We do not elaborate on other aspects of information disorder, such as malinformation, in the context of this report. In our view, disinformation is the more common term used to describe intentional misinformation. For these reasons, we use the umbrella term of misinformation throughout the report to refer to both misinformation and disinformation.</p>
Fake news	<p>The New York Times (2016) defines “fake news” as deliberately fabricated information used as clickbait to generate internet traffic for profit. Clickbait is content whose primary purpose is “to attract attention and encourage visitors to click on a particular webpage link” (The Oxford English Dictionary in The Atlantic, 2014, webpage).</p>
Infodemic	<p>According to the WHO (2020) an infodemic is:</p> <p>...an overabundance of information, both online and offline. It includes deliberate attempts to disseminate wrong information to undermine the public health response and advance alternative agendas of groups or individuals. Mis- and disinformation can be harmful to people’s physical and mental health; increase stigmatization; threaten precious health gains; and lead to poor observance of public health measures, thus reducing their effectiveness and endangering countries’ ability to stop the pandemic (webpage).</p>
Hate speech	<p>Mondal, Silva, and Benevenuto (2008) identify hate speech “as an offensive post, motivated, in whole or in part, by the writer’s bias against an aspect of a group of people” (p. 87).</p>



Machine learning	Amazon (2021) states that: Machine learning (ML) and deep learning (DL) are both computer science fields derived from the discipline of Artificial Intelligence. Broadly, these techniques are separated into “supervised” and “unsupervised” learning techniques, where “supervised” uses training data that includes the desired output, and “unsupervised” uses training data without the desired output (webpage).
Misinformation	See definition of <i>disinformation</i> above.

Table 2

¹This list is not exhaustive due to the limited scope of the sources consulted, all of which were published in the English language.





Section 3: Main Findings – Part 1

Identification of misinformation

This section begins by mapping out the phenomenon of misinformation. As a modern and growing phenomenon, it is important to conceptualize misinformation and what we mean by it. First, we look at the features and identification of online misinformation. Online misinformation is different from the offline misinformation found in daily conversation and traditional print media due to several critical factors, including the breadth and design of misinformation spread, the massive variety of information sources, and its profitability.

Second, the effects of misinformation are still not well understood. We highlight this point because research has not been able to keep up with developments around misinformation and how to study this rapidly changing modern phenomenon. Third, we discuss how populations are vulnerable to, or targeted for, misinformation campaigns, and how they affect political processes, elections, international cooperation, and democracy.

We complete this section by looking at the relationship between misinformation, hate speech, and physical violence, including mass atrocities. A portion of this discussion looks at the link between viruses and violence in light of the COVID-19 pandemic. We conclude with a consideration of the tools that have increased online to reduce misinformation. This background study ultimately sets us up to answer the timely question “How can misinformation management effectively and sustainably operate at scales involving mass data quantities and audiences which cannot be engaged on an interpersonal level?”

This research reveals several important themes that need to be considered when understanding the breadth of the phenomenon of misinformation. In this section, we will discuss the following:

- The distinction between online and offline misinformation
 - The speed, breadth, and design of misinformation
 - Variety of sources
 - Profitability
 - Demographic factors
- How populations are targeted for misinformation campaigns
 - Politics and elections
 - International cooperation efforts
- The relationship between misinformation, hate speech, mass atrocities, and impacts upon democracy
 - Violence and viruses

The distinction between online and offline misinformation

Several studies have identified how online misinformation is distinct from offline misinformation in at least five essential ways:

1. Online misinformation has the potential to affect people on a scale far more expansive and at a faster rate than offline misinformation



2. The variety of information has increased
3. Online misinformation is profitable
4. Specific populations are more vulnerable and targeted by misinformation campaigns
5. Researchers have not been able to keep up with understanding to what extent online misinformation is affecting people's decision-making

The creation, speed, breadth, and design of misinformation spread

One of the critical distinctions of online misinformation is the speed of its reach. In a mapping study of online misinformation titled *The Science Behind Fake News*, researchers Lazer et al. (2016) argue in an interview on the topic that:

We just don't have time to separate the facts from the falsities. Even fact-checkers don't have time. A message can go viral before any serious truth filter has been applied. This leads to a positive feedback cycle. If one team is passing around unfiltered information then its opponents feel the need to respond in real time. So they won't be able to check their facts either. The positive feedback is further increased by individuals' tendency not to read very carefully. People share messages without even reading them, never mind evaluating their accuracy. (webpage)

According to a recent study conducted by MIT researchers in 2018, researchers observed that:

Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. (Vosoughi, 2018, p. 359)

A BBC News report on this study of 126,000 rumours and false news stories indicated that "they [rumours and false news stories] travelled faster and reached more people than the truth" (webpage). Indeed, a 2017 Brookings report pointed out that a significant difference between offline and online media is that "to maximize reach, traditional outlets curated information for veracity and balance. In stark contrast, the curation of social media platforms is not for veracity, but for advertising velocity" (webpage). On social media platforms where individuals can message one another directly, there have been significant issues with the rapid forwarding of misinformation. Related to the COVID-19 pandemic, in a joint press release, the WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC has called the misinformation around COVID-19 an "infodemic" (WHO, 23 September 2020, webpage). These examples show how misinformation proliferates much faster than misinformation shared through traditional print media.

Online misinformation is also potentially destructive because of technology's ability to effectively and efficiently dial into human emotional responses. Lazer et al. (2018) further show that misinformation on Twitter, for example, is typically retweeted by many more people and far more rapidly than facts. Indeed, the same Brookings article explained that "the software algorithms that decide our news feed are programmed to prioritize user attention over truth to optimize for engagement, which means optimizing for outrage, anger, and awe" (webpage). As Wheeler (2017) (writing for Brookings) elaborates further, the longer a social media platform can hold a user's attention, the greater the platform's economic success. In order to capture audiences, social media companies gather information about the user so that the user could be later targeted with content that they are likely to enjoy, agree with, or want to buy. In another study, the Pew Research Center and Elon University's Imagining the Internet Center canvassed subject matter experts to ask what they thought the misinformation landscape would look like in a decade. Of the 51% of participants who believed that things would not improve generally, one of the two main reasons was because:



...humans' primal quest for success and power – their “survival” instinct – will continue to degrade the online information environment in the next decade. They predicted that manipulative actors will use new digital tools to take advantage of humans' inbred preference for comfort and convenience and their craving for the answers they find in reinforcing echo chambers.

Furthermore, the internal logic that drives people towards specific content is a novel instrument that users may not be aware of. Although the users may think they are intentionally reading something out of interest, they would not know that they had already been profiled and primed to read specific articles. One recent Princeton study, published in *Nature*, found that Facebook was the platform with the most prolific misinformation spread. Studying the prevalence of misinformation leading up to the 2016 American elections, the researchers observed that Facebook referred users to untrustworthy news sources over 15% of the time. At the same time, another scholar argues that Twitter does less to reduce misinformation than other social media sites. While recognizing the steps Facebook has implemented, in a 2017 opinion editorial, he stated that:

Twitter has done the opposite—its security team is rudimentary and reclusive; the company seems to be in denial on the scope of disinformation; and it even optimised its platform for hiding bots and helping adversarial operators to delete incriminating evidence—to delete incriminating evidence not just from Twitter, but even from the archives of third party data providers.

Lastly, Lazer et al. (2018) point out that bots also contribute to the spread of disinformation. By liking, sharing, and searching for information, bots (in their study, on Facebook) magnify the spread of misinformation by speeds impossible for individual humans to achieve. A 2020 report in *The Atlantic* shared this concern, stating that artificial intelligence (AI) will create an infinite supply of misinformation. The author points out that, for example, a “tool called GPT-3 generates long-form

articles as effortlessly as it composes tweets, and its output is often difficult to distinguish from the work of human beings.” The author, Renée DiResta, of the Stanford Internet Observatory, continues:

In countries around the world, coordinated propaganda campaigns in print as well as social media have sown social unrest, pushed down vaccination rates, and even promoted ethnic violence. Now imagine what happens when the sources of such postings are untraceable and the supply is essentially infinite.

The variety of sources of information

In comparison to traditional print media, the massive range of sources is another characteristic of online misinformation. In the same interview on the research report, *The Science of Fake News*, one of the researchers quoted above explained that:

...because of technology, anyone in the world can be a source of news... the huge variety of news media in our culture means that people have the freedom to tune into news sources that tell them what they want to hear and we all like to hear news consistent with our beliefs.

A recent CBC article showed that, according to a study by the University of Sherbrooke, one in ten Canadians believes in a COVID-19 conspiracy theory. In another recent survey, CBC found that one in four Canadians believes that government warnings about COVID-19 are out of proportion to the actual threat (CBC News, 2020). In a BBC article, Kevin Kelly, the co-founder of *Wired* magazine, attempted to explain the increase in misinformation. Kelly noted that it could be partly because “truth is no longer dictated by authorities, but is networked by peers. For every fact, there is a counter-fact, and all these counter-facts and facts look identical online, which is confusing to most people.” This vast array of sources of news and information makes it more difficult to discern between genuine information and misinformation.



Misinformation is profitable

Individuals and organizations often traffic in misinformation as a source of income. One well-known article on Wired (2017) explained how young people in a small town in Macedonia called Veles published misinformation for profit since they generated income from hosting advertisements alongside their fabricated news stories. Another article by the BBC (2018) discussed how one man in the United States, coined “the Godfather of Fake News,” published misinformation for profit. Indeed, misinformation can make individuals as well as companies quite wealthy. As one 2016 article in The Guardian points out, “what he [Mark Zuckerberg] omits to mention is that Facebook has a conflict of interest in these matters. It makes its vast living, remember, from monitoring and making money from the data trails of its users. The more something is ‘shared’ on the Internet, the more lucrative it is for Facebook.” A report from a workshop organized by the Canadian Security Intelligence Service (CSIS) in 2018² asserted that misinformation actors:

...profit from click-based advertising directed at readers of sensationalist stories and those who limit their news consumption to online news aggregating web sites. These enterprises maximise their readership and clickbait potential by purchasing the pages of groups with sizeable memberships which fit the target demographic. The truth, falsehood, and subject matter of their news content are irrelevant—the singular objective is attracting readers who will view advertisements. (p. 99)

Misinformation is also inexpensive to create and is therefore a cheap tool for fomenting discord. As Shu et

al. (2020) point out “Thanks to the low cost of creating fake news sources and the software-controlled social media bots, it has never been easier to shape public opinion for political or financial reasons” (p. 8). The MIT Technology Review also reported in 2017 on the low cost of producing misinformation. It pulled findings from a report by TrendMicro which calculated that misinformation could be used to discredit a journalist for \$55,000 or to incite a protest for \$200,000. The MIT Technology review stated that:

...as long as it remains so cheap to shape public perception using fake content—and these figures are, after all, peanuts compared with the advertising budgets behind plenty of real content—they [social media companies] have a fight on their hands. (webpage).

²As is written in the report, “This report is based on the views expressed during, and short papers contributed by speakers at, a workshop organised by the Canadian Security Intelligence Service as part of its academic outreach program. Offered as a means to support ongoing discussion, the report does not constitute an analytical document, nor does it represent any formal position of the organisations involved. The workshop was conducted under the Chatham House rule; therefore no attributions are made and the identity of speakers and participants is not disclosed.”



The effects of misinformation are not well understood

Research has yet to ascertain to what degree misinformation is impacting people's decision-making capacities. An early study by Betsch (2011) looked at the effects of access to online information regarding an individual's decision-making process related to health. He observed that data from internet searches could influence an individual's decision-making in ways that they themselves did not yet consciously perceive (Betsch, 2011). The Internet Health Report by Mozilla (2019) pointed out that the topic recommendation algorithms used by many companies actually promote misinformation about various issues, including vaccines, cancer, gender discrimination, terrorism, cult, and conspiracy theories. Algorithms suggest new content to users in order to propel engagement and lead people to click on misinformation (Mozilla, 2019).

Simultaneously, the study by Lazer et al. (2018) which is mentioned above points out that evaluations of impacts on political behaviour, such as exposure to misinformation, are virtually non-existent in the literature. Nevertheless, we see that populations are targeted because of who they are. We also see that people are especially targeted during politically volatile periods, such as elections, and that some actors want to cause disruption through misinformation that impacts social relations, sometimes leading to violence and even mass atrocities.

How populations are targeted for misinformation campaigns

Numerous studies have identified how populations are targeted for disinformation campaigns in several distinct ways, especially looking at how specific demographic groups are chosen as targets in these cases. Studies also mainly centre on politics, elections, and how it

impacts elections and democracy more broadly. This section ends with a consideration of misinformation and epidemics and pandemics, including the current COVID-19 pandemic.

Demographic factors

Specific populations are more or less susceptible to consuming misinformation and demographic factors also impact this consumption. For example, with age, a study by the Pew Research Centre found that young people are less likely to fall for misinformation than older people, stating that "younger adults are still better than their elders at deciphering factual from opinion news statements" (Gottfried & Grieco, 2018, webpage). Another joint study by Princeton University and New York University (2019) discovered that 65-year-olds are much more likely to share misinformation on Facebook than 18 – 29-year-olds. Looking at the connection between political beliefs and misinformation consumption, researchers in another study uncovered that liberals and conservatives are similarly susceptible to fake news. The study observes that people on both ends of the political spectrum are likely to believe news that is consistent with their ideology and disbelieve information that is inconsistent with it (Harper & Baguley, 2019).

A similar study from 2016 had similar results (Taber & Lodge, 2016). As the authors write, "our studies show that people are often unable to escape the pull of their prior attitudes and beliefs, which guide the processing of new information in a predictable and sometimes insidious way" (p. 767). A more recent study from 2019 determined that the level of polarization in a given society affects people's susceptibility to misinformation. The authors, Vicario, Quattrociocchi, Scala, and Zollo (2019), state that "Users' polarization and confirmation bias play a key role in misinformation spreading on online social media" (p. 1). Knowing this, the authors develop a framework to proactively determine potential misinformation targets and aim to identify polarizing content on social media before they become full-blown misinformation topics (Vicario, Quattrociocchi, Scala, & Zollo, 2019). It would be interesting to learn if the effects that they found hold in



other contexts, such as in countries in the Global South.

The recent 2019 State Department report charted the psychological factors that drive vulnerabilities to misinformation. Psychological factors include emotions and biases, the brain's mode of simplifying complex topics, the human proclivity to want to expose truths, and the feeling of validating one's identity. Shu et al. (2020) give the example of Bronstein et al. (2019), who studied the reasons that might lead people to believe in fake news. Bronstein et al. (2019) label the reasons as delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. According to them, people who hold delusion-like beliefs, dogmatic individuals, or religious fundamentalists are more likely to believe misinformation because they are less likely to engage in open-minded thinking and analytical thinking (Bronstein et al., 2019). Shu et al. (2020) highlight how Pennycook and Rand (2019a, 2019b) discovered that accepting misinformation at face value is more related to lack of thought than to ideological or partisan bias.

In an interview on the topic, Pennycook stated that "People in our study did not act like intense partisans in the context of fake news. Rather, those who fell for fake news were those who were just being lazy cognitively. A bit more effort might go a long way" (webpage). Finally, Shu et al., (2020) outline that, according to Bovet and Makse (2019), "confirmation bias and social influence is the main reason for echo chambers where users with similar beliefs share disinformation about a specific topic." These studies demonstrate the links between misinformation and demographic/psychological factors that drive people to be more vulnerable to it. It would be valuable to have more research findings that relate to people's tendencies to believe misinformation in conflict-prone or historically conflict-ridden environments which are also information deprived. In these specific contexts, people may be driven to more readily believe misinformation because of the constant need to be alert to potential violence.

National politics and elections

Misinformation is prevalent in politics, especially during elections. For example, a US State Department report looks at foreign state-sponsored misinformation originating in Russia, China, Iran, and North Korea against other countries, companies, and citizens. Misinformation (in this case disinformation) is manufactured to sow disruptions and chaos in the hope of achieving certain political ends. In both the Global North and the Global South, elections have become a central focal point for disinformation campaigns. The same State Department report pointed out that:

As the adoption of new technology and social media platforms have spread globally, so too have government efforts to exploit these platforms for their own interests, at home and abroad. Russian attempts to influence the United States 2016 presidential election and the 2016 Brexit vote in the United Kingdom are two recent, high-profile examples (p. 14).

A lot of research has culminated around the 2016 and 2020 elections in the United States. A study by Allcott and Gentzkow (2017) estimated that "the average U.S. adult read and remembered on the order of one or perhaps several fake news articles during the election period, with higher exposure to pro-Trump articles than pro-Clinton articles" (p. 232). Another report by Bovet and Makse (2017) in Nature used a dataset of 171 million tweets from the five months preceding the election day to understand the influence of misinformation during elections. They determined that:

...25% of these tweets spread either fake or extremely biased news...we find that, while top influencers spreading traditional center and left-leaning news largely influence the activity of Clinton supporters, this causality is reversed for the fake news: the activity of Trump supporters influences the dynamics of the top fake news spreaders (p. 1).



In addition to misinformation affecting specific demographics, some populations are deliberately made into misinformation targets. For example, NPR recounted in 2017 that when FBI agent Clint Watts testified before the Senate Intelligence Committee, he described how Russian Twitter bots spread misinformation to target potential Republican voters in the midwestern United States. Similarly, a Washington Post article from 2018 explained how “the Kremlin engaged in a coordinated campaign to elevate Donald Trump to the presidency, and this country’s technology companies were central to its strategy” (webpage). Although the impacts of such campaigns are still being studied, these examples demonstrate how misinformation over social media is a growing source of concern, especially during elections. Another study of Russian-led misinformation campaigns during the 2020 American elections stated that:

Targeted disinformation campaigns are a cheap and effective way to create real harms that have a society-wide impact. This form of information warfare capitalizes on inherent features of the internet messaging platforms and the free nature of democratic societies to spread false and malicious content designed to increase discord by exacerbating existing social and political chasms, promote chaos and fear, and generate distrust toward government (p. 2).

There have been reports of the Russian-led misinformation campaigns used to foment racial tensions and suppress Black citizens’ votes in the United States during the 2016 elections. Again in 2020, the same tactics were used by Russian-led campaigns and others, including American-led misinformation campaigns against their own elections. An additional news report found that Asian-Americans and Latino communities are targeted through social media platforms like WeChat, WhatsApp, Facebook, KakaoTalk and YouTube to consume right-wing conspiracy theories. The article stated that older immigrants who use these platforms but who do not usually watch mainstream media in English are susceptible to misinformation about American politics through these channels. In particular,

the news report notes that since misinformation about the Democratic Party being socialist has circulated “older Korean, Chinese, Taiwanese and Vietnamese immigrants who fear anything left-of-center” might believe that the Democratic Party is becoming communist. Another article in The Atlantic showed how Russian intelligence agents create fake personas to distribute misinformation about a variety of themes including, “Syria, Black Lives Matter, and Hillary Clinton’s emails” (webpage).

Misinformation during elections goes beyond the United States. In Canada, in 2019, a CBC/Radio-Canada analysis of 9.6 million tweets showed that foreign Twitter trolls emboldened conspiracy theorists in the lead-up to the 2018 federal election over sensitive political issues in the country. In response to such threats, the Canadian government set up the Security and Intelligence Threats to Elections Task Force to counter hostile actors who attempt to interfere with electoral processes. Cases of misinformation during elections have also impacted European elections. A Toronto Star article describes an information leak against France’s incumbent president during the country’s 2017 presidential election. The article quotes a statement from Macron’s team saying that the documents had been mixed with false documents to “seed doubt and disinformation” in order to undermine the presidential runoff vote. Although not on social media, this example of leaked documents highlights the link between misinformation and efforts to destabilize elections. The Financial Times reported about how misinformation related to refugee policy impacted Germany’s 2017 elections. In one example, Renate Künast, a member of parliament for the Green Party, was a misinformation target and she has since sued the Resistance of German Patriots, a right-wing group, for its involvement. German chancellor Angela Merkel has also been targeted by misinformation campaigns, mostly directed at her open-door policy toward refugees.

In another case, a recent article in ICTworks described how in Nigeria an Israeli company called the Archimedes Group spent \$812,000 on Facebook to promote pages, groups, events, and Instagram accounts. Due to this



activity, Facebook revealed that the network grew and that many of the Archimedes Group's fake Facebook accounts represented themselves as Nigerian. The article cited the Times of Israel, which indicated that the views in the shared information shared the goal of reelecting Nigerian president Muhammadu Buhari. In another political example, an Al Jazeera online interview analyzes the links between incumbent Kenyan president Uhuru Kenyatta and the international data firm Cambridge Analytica. In an article from 2017, the Red Cross in Kenya was concerned about violence related to the election and post-poll protests derived from misinformation. The Red Cross urged Kenyans, "especially those on social media, to exercise restraint and verify every information and photos before sharing them."

BBC reported in 2019 on how the popular messaging platform WhatsApp had become a vehicle for misinformation in India, especially ahead of the 2019 federal election. Similarly, another BBC article from 2018 showed that nationalism drove people to share misinformation related to politics in the country. Australia and Denmark set up task forces in 2018 to monitor possible cyber attacks based on misinformation promoted by other state actors and foreign sources during their elections. In 2018, The Guardian reported on the UK government's announcement that its National Security Communications Team (NSCT) would start countering misinformation distributed by state actors. In France, the National Commission for the Control of the Electoral Campaign for the Presidential Election and the National Cybersecurity Agency countered Russian efforts to interfere in the country's 2017 presidential election.

Regarding COVID-19, the United Nations Department of Global Communications (DGC) issued several statements on how the pandemic has proliferated misinformation around the virus. They stated that the World Health Organization (WHO) had put together a group of "mythbusters" working with search and media companies to counter the spread of rumours. As the President of the United Nations General Assembly (UNGA) calls for a multilateral response to COVID-19, some experts feared that misinformation regarding the origins of the virus

could have hampered international cooperation on the pandemic, particularly with Chinese scientists who are often the subjects of misinformation. Simultaneously, the Chinese government "has undertaken numerous measures to change the COVID-19 narrative and disassociate itself from COVID-19" (Verma, 2020, p. 248).

The effects of misinformation on democracy

The report of the High-Level Expert Group on Fake News and Online Disinformation (HLEG) (2019) states that disinformation can be harmful to citizens and democratic societies at large. Threats stemming from misinformation include those related to democratic political processes, including the integrity of elections and democratic values that shape public policies across sectors (health, science, finance, etc.). In Canada in 2018, The Toronto Star reported that CSIS had identified a broad range of misinformation actors including "a 20-something in Kosovo pushing out 'fake news' to make a quick buck, to hyper-partisans trying to influence domestic politics, to sophisticated influence campaigns from hostile nations trying to exploit existing divisions in Western society." Again in 2020, Global News indicated that CSIS had determined that Russia, China, and Iran were spreading misinformation related to COVID-19 to promote their political objectives. In the same article, Global News quoted CSIS spokesman John Townsend, who asserted that:

It is important to note that disinformation—originating from anywhere in the world—can have serious consequences, including threats to the safety and security of Canadians, erosion of trust in our democratic institutions, and confusion about government policies and notices— including information on the COVID-19 pandemic.



In another example related to Brexit, a 2018 CSIS workshop report noted that:

While the content tweeted on Brexit has a stronger slant towards nationalist and nativist values compared to the content tweeted by the global population (27 per cent versus 19 per cent, respectively), the emerging reality of hyper-partisan websites is that they cater to both extremes of the political spectrum, are often owned by the same companies, and repurpose stories to accommodate and confirm readership bias (p. 58).

The relationship between misinformation, hate speech, and atrocities

The heart of this report, which is also a large part of the SP's work on the ground, is related to understanding the links between misinformation, hate speech, and mass atrocities. Several studies and research reports have examined the relationship between misinformation, hate speech, and physical violence—including mass atrocities. Some of the most notable examples of such incidents from recent years are described in the table below.

India

In 2017, the New York Times reported that mob violence in India had been incited by misinformation circulated on WhatsApp, referring to examples of mob violence in Mumbai, Delhi, and in the states of Rajasthan and Uttar Pradesh. A 2018 BBC report also described a mob attack on three men visiting relatives in Handikera. The three men became the targets of violence after a message about potential child abductors was circulated on WhatsApp and one of the men was killed. Another BBC report describes mob lynchings that occurred throughout India in 2018 which were also linked to misinformation circulated on WhatsApp. Similarly, in a study of lynching in India, Chinmayi (2019) concludes that these incidents are more related to incitement to violence than solely misinformation proliferated online. He adds nuance to the problem by stating that rumour verification is only one part of the equation, pointing out that research needs to be conducted at the institutional level to know why hateful misinformation exists in the first place and how to target those root causes. He states that:

Little has been done by political leaders to mitigate the sense of insecurity and entitlement that has led to the violence. It would take a detailed study of the incidents of violence to fully comprehend the combination of factors that leads to and sustains the violence.



Sri Lanka

A detailed analysis of violence in Sri Lanka by the New York Times (2018) shows how misinformation, spread through Facebook, encouraged violence between Muslim Tamil people and Buddhist Sinhalese people in the country. For instance, a rumour shared on Facebook about a restaurant owner putting sterilization pills into the food that he sells led to him being beaten, his shop destroyed, and a local mosque being burned down. Another study on violence in Sri Lanka by the Toda Peace Institute found that “today, violent content on social media is often the digital manifestation of longer-standing communal fears, anxieties and concerns. These socio-political tensions have now metastasised into short-form video, memes and tweets produced by and for a young demographic” (p. 2). According to this report, social media is not the cause of violence, but it is a new vehicle for misinformation that speeds up the delivery of hateful and violence-inciting messages.

South Sudan

South Sudan has been the site of a civil war and related rebel activity since 2013. Referring to renewed clashes in Juba—the capital and largest city of South Sudan—a 2016 UN Report described how “social media has been used by partisans on all sides, including some senior government officials, to exaggerate incidents, spread falsehoods and veiled threats or post outright messages of incitement” (p. 10). UN News also stated in 2016 that “The media, including social media, are being used to spread hatred and encourage ethnic polarization.” #Defyhatenow, an initiative that works on providing community-based and data-driven solutions to the problem of hate speech, misinformation, and disinformation, wrote that:

...even in a country where only a small fraction of citizens are formally educated or have access to the Internet, social media and the language of hate, incitement and directed online attacks have had the ability to stoke conflict, exacerbating attempts at rebuilding trust and dialogue amongst communities. (webpage)

Another company, PeaceTech Lab, points out that online hate speech was a concern even before 2013 and that diaspora communities often use inflammatory language and images to voice their grievances. In order to better understand the connection between hateful narratives online and violence on the ground in South Sudan, they published a lexicon in 2016 to support efforts to identify hate speech in the country

Nigeria

In a BBC report on misinformation in Nigeria, the reporters show that violence between the Fulani and Berom ethnic groups has been exacerbated by online misinformation. The article states that:

In Nigeria, the lines between disinformation and hate speech are often blurred. What is clear is that, despite Facebook’s attempts to root out hate speech, and the efforts they are making to detect and remove false information, hundreds of inflammatory posts are slipping through the net.

Another 2020 BBC report explained how misinformation about armed robbers in Ogun, Nigeria caused people to panic and brace for violence. The rumours spread on Twitter, with the hashtags #OgunUnrest and #LagosUnrest. However, as the article notes, the police did not find any evidence of armed robbery, pointing to the explosive consequences of misinformation that have no standing in on-the-ground realities.



Mexico

In Acatlán, Mexico, Ricardo and Alberto Flores were burned to death in 2018 after a mob of people mistakenly identified them as child abductors who had been rumoured on WhatsApp to be operating in the same area.

Myanmar

Particular attention should be paid to how the Myanmar military has used misinformation to facilitate genocide against Rohingyas in Rakhine State in western Myanmar. In 2013, SP published a report about the risk of genocide in Myanmar, citing, among ten other reasons, “the official denial of the existence of the Rohingya and institutionalized usage of hate speech even among moderates” (p. 4). The report further points out that the rejection of the term “Rohingya” and the use of the government’s preferred term “Bengali” in reference to people who identify as “Rohingya” is a policy that “constitutes hate speech because it implies that they are illegal immigrants” (p. 5). SP points out that “The government officially refers to the Rohingya as “Bengalis,” a tactful cover for hate speech used to exclude them from the many officially recognized ethnicities of Burma” (p. 32). Indeed, Rohingyas are not able to obtain official citizenship, barring them from all associated rights. SP additionally finds that the term has been used by state media, along with the term “kalar,” which is a “racial slur for any person of Muslim or Indian appearance” (p. 60). The report explains that state and state-sanctioned actors spread hate speech through traditional and social media. One particularly egregious example is the hate speech propagated by the Buddhist nationalist monk Wirathu. The report explains that:

Unverified and often unfounded rumours play a large part in catalyzing ethnic violence and genocide, with Burma’s violence being no exception. Wirathu has been central in recycling some longstanding rumours and myths about the Rohingya while also fabricating and disseminating new ones. He alleges, for instance, that Muslims commit 100 per cent of rape cases in Burma and that intermarriage is forced upon Burmese to increase the Muslim birthrate and outnumber Buddhists (Muslims represent a mere four to five per cent of Burma’s population).³ He also alleges an international conspiracy of economic attacks on Buddhist businesses financed by Saudi oil money⁴ and that Muslims burned down their own houses in Meikhtila so that they could receive international aid.⁵

In another example, the 2018 report by the United Nations Fact-Finding Mission in Myanmar noted how the country’s commander-in-chief, Senior General Min Aung Hlaing, used Facebook for strategic political ends. On 2 September 2018, referring to Rohingyas, he stated that “the Bengali problem was a longstanding one which has become an unfinished job despite the efforts of the previous governments to solve it. The government in office is taking great care in solving the problem” (p. 8). The report goes on to describe how “The Myanmar authorities, including the government and the Tatmadaw, have fostered a climate in which hate speech thrives, human rights violations are legitimized, and incitement to discrimination and violence facilitated” (p. 14). Although not a central focus of the report, the authors note the importance of social media in fostering violence. They write that:

³Wade, F. (2013, April 22). The monks who Hate Muslims. Retrieved March 01, 2021, from <https://foreignpolicy.com/2013/04/22/the-monks-who-hate-muslims>

⁴“Myanmar’s ‘969’ Crusade breeds anti-Muslim malice.” (2013, March 27). Retrieved March 01, 2021, from <https://www.pri.org/stories/2013-03-27/myanmar-s-969-crusade-breeds-anti-muslim-malice>

⁵“Man of peace’ Venerable Wirathu approves violence to Rohingyas.” (n.d.). Retrieved March 1, 2021, from <https://www.theaustralian.com.au/news/world/man-of-peace-venerable-wirathu-approves-violence-to-rohingyas/news-story/79bc44541116c9662191e9769dbba8b3>



The role of social media is significant. Facebook has been a useful instrument for those seeking to spread hate, in a context where, for most users, Facebook is the Internet. Although improved in recent months, the response of Facebook has been slow and ineffective. The extent to which Facebook posts and messages have led to real-world discrimination and violence must be independently and thoroughly examined. The mission regrets that Facebook is unable to provide country-specific data about the spread of hate speech on its platform, which is imperative to assess the adequacy of its response (p. 14).

Shortly after the IIFFMM report was published, Facebook contracted the Business for Social Responsibility (BSR) to conduct a human rights impact assessment of Facebook's services in Myanmar. Facebook published BSR's findings in November 2018, acknowledging that "we weren't doing enough to help prevent our platform from being used to foment division and incite offline violence".

There have been several studies on the links between misinformation, hate speech, and violence in Myanmar. Researchers conducting joint research for Save the Children and the University of Sydney interviewed and ran focus groups with youths in Rakhine and Kayah states, which are ethnically diverse. The researchers determined that Rohingyas, who are mainly Muslims, were often targets of online hate speech and misinformation, followed by Christians. The researchers note that Maramgyi youths were also discriminated against and called "kalar" by the Rakhine community. They further point out that Chin youths also described experiences of marginalization by the Rakhine community, who in turn face discrimination by the Bamar community. In response to these findings, the researchers state that "the overlap between fake news and hate speech means that encouraging culture change around tolerance and acceptance of people from different ethnicities (particularly Muslims who were the most common

target of hate speech in the current study) will also be important going forward" (p. 101).

Beyond reports by international organizations, news outlets have also reported on the treatment of Rohingyas on social media in Myanmar. In 2018, the Financial Times wrote about online misinformation and the freedom of the press in Myanmar. The article outlined how, after the August 2017 state-led attacks against Rohingyas in Rakhine State, "twin chain messages were widely distributed via Facebook Messenger.

Violence and viruses

The relationship between misinformation and viruses existed before the COVID-19 pandemic. For example, in India rumours that spread over WhatsApp hindered a measles-rubella vaccination campaign because people heard that it caused children to become sick. In another example, during the Ebola epidemic in the Democratic Republic of the Congo (DRC), a study by the Harvard Humanitarian Initiative determined that 13% of WhatsApp messages over a 34-day period across eight large group chats referenced or spread rumours and misinformation about the Ebola virus. A Time magazine article indicated that:

...misinformation has contributed to the difficulties containing the virus in DRC, where more than 1,300 people have now died in the second-largest Ebola outbreak in recorded history. Rumours have led local Congolese to distrust both medical workers and the authorities' response.

According to the article, some misinformation refers to the central government's alleged role in spreading the virus. In contrast, other misinformation narratives refer to the idea that international aid organizations are responsible for the outbreaks. The effects of misinformation became deadly in the case of the Ebola virus. Indeed, the New Humanitarian reported that "Beating back Ebola required a pitched battle not just against the disease, but also against lies, conspiracy



theories, and attacks on public health workers.” Medecins Sans Frontieres announced that they had to discontinue operations during the epidemic’s height in order to protect healthcare workers in medical facilities from physical attacks.

The problem of misinformation spreading about COVID-19 has been even more acute. In the early days of the pandemic (February 2020) BBC News reported that Chinese tourists in Ukraine were attacked by local people who believed that they had COVID-19. The BBC article stated that “Ukraine’s security service (SBU) said a fake email claiming to be from the health ministry falsely said some evacuees had contracted the virus.”

Early on in the pandemic, in Hungary, Human Rights Watch (HRW) noted that the government used COVID-19 as a pretense for xenophobic and anti-immigrant sentiment. Hungarian president Viktor Orbán said at a news conference that “We observe a certain link between coronavirus and illegal migrants” (Euro News, 2020). In a transnational case, Reuters found evidence in a European Union (EU) document that the EU believed that Russia had orchestrated an online misinformation campaign related to COVID-19 targeting European democracies. The misinformation, released “in English, Spanish, Italian, German, and French, uses contradictory, confusing, and malicious reports to make it harder for the EU to communicate its response to the pandemic.”

The International Rescue Committee (IRC) asserted that they had noticed a surge in misinformation, specifically regarding manipulating health messaging. For example, they point out that misinformation is used for short-term political gain in Colombia, making people believe that the virus “is a conspiracy created by the governments to divert attention from issues such as corruption, violence, unemployment, and social inequality.”

Similarly, the Associated Press reported that Yemen, which is already a conflict zone, also experienced rumours related to COVID-19 that impacted how the population

dealt with the virus. One particularly pernicious rumour was that “Houthi rebels have instructed doctors to kill suspected COVID-19 patients with a ‘mercy injection.’” The Associated Press described how Houthi leaders denounced the rumour but the article pointed out that Houthis also spread rumours that foreigners spread the virus. The article concludes that rumours caused panic and made people less willing to inform health officials about potential cases of COVID-19. Misinformation related to COVID-19 has also been prevalent in Canada. The CBC reported in March 2020 that:

Gen. Jonathan Vance, the country’s chief of the defence staff, said he’s seen indications recently that Canada’s adversaries intend to exploit the uncertainty, confusion and fear that many people feel after a week marked by swift and extraordinary developments in the global pandemic crisis.

These articles show that countries worldwide have been vulnerable to misinformation campaigns, both deliberate and inadvertent, that have hampered COVID-19 containment efforts.

Misinformation verification and establishing facts – What do we do to counter misinformation?

This section of the report was developed in response to the question “How can misinformation management effectively and sustainably operate at scales involving mass data quantities and audiences which cannot be engaged on an interpersonal level?” It addresses the techniques that have been developed by the public and private sectors to verify and establish facts in order to counter misinformation.



Examples of joint content flagging initiatives

Shu et al. (2020) highlight that a common way to counter misinformation on social media is to flag and fact check posts. Generally, the standard practice of social media companies is that if a post is frequently flagged then it is fact checked, after which it is removed from the platform if it is deemed to be misinformation. Shu et al. (2020) also point out that many initiatives are geared towards source identification. At the same time, figuring out where sources originate or are distributed from is an important way to counter misinformation. They further point out that “identifying these ‘opinion leaders’ propagating disinformation and terminating their accounts or even slowing down their reach may slow down the spread of disinformation” (p. 21). The sections below give examples of the types of initiatives that flag and track down misinformation spreaders.

The Network Contagion Research Institute (NCRI) is a team bringing together experts from private industry, academia, and the public sector who have developed Contextus, a platform used to trace and uncover deception and hate on social media. In another example, as part of their Truth Decay project, the RAND Corporation published an extensive database of tools for fact-checking. Truth Decay, which is a term coined by RAND, is defined by:

- Increasing disagreement about facts and analytical interpretations of facts and data
- A blurring of the line between opinion and fact
- Increasing relative volume and resulting influence of opinion and personal experience over fact
- Declining trust in formerly respected sources of facts

The International Fact-Checking Network (IFCN), which is a part of the Poynter Institute, is an umbrella group of fact checkers which regulates and certifies fact checking organizations. Driven by the idea that transparent fact checking can be powerful, the IFCN developed a code that organizations commit to in order to promote a

culture of fact checking. In another example, PolitiFact has developed the Truth-O-Meter methodology for independent fact checking. This initiative sees editors and reporters jointly rating the accuracy of political statements based on a 6-point scale.

Examples of universities, public institutions, news media, social media companies

Many universities and public institutions have produced tools for identifying misinformation, such as the Toronto Public Library's list of questions a person should ask themselves before trusting a source. This resource includes a list of fact checking resources. The University of Toronto Library also published a list of resources guiding people on how to spot misinformation. NPR developed a five-point checklist to help people determine whether they are reading misinformation. Cornell University Library put together an eight-point infographic to help readers evaluate a piece of information that they are unsure about. These points are as follows:

- Consider the Source – Click away from the story to investigate the site, its mission, and its contact info.
- Read Beyond – Headlines can be outrageous in an effort to get clicks. What's the whole story?
- Check the Author – Do a quick search on the author. Are they credible? Are they real?
- Supporting Sources? – Click on those links. Determine if the info given actually supports the story.



- Check the Date – Reposting old news stories doesn't mean they're relevant to current events.
- Is it a Joke? – If it is too outlandish, it might be satire. Research the site and author to be sure.
- Check your Biases – Consider if your own beliefs could affect your judgment.
- Ask the Experts – Ask a librarian or consult a fact-checking site.

CBC also compiled a question guide to help readers verify sources, and their own list of resources, which include the following:

- FactsCan.ca
- Trudeaumetre
- FactCheck.org
- PolitiFact.com
- The Washington Post's Fact Checker
- Poynter's Fact-Checking page
- FullFact.org in the UK
- Verafiles.org in the Philippines
- Africacheck.org

One large-scale initiative is Facebook's Fact-checking initiative involving third-party fact checking organizations which are certified through the IFCN in order to identify, review, and counter misinformation. The program involves first identifying misinformation and sending it to fact checkers (Facebook Fact-checking, 2021) who then review the content, check its facts, and rate its accuracy. The fact-checkers label the misinformation as applicable and inform users of its falsity (Facebook Fact-checking, 2021). Once it is verified, in order to reduce its visibility, the false content then becomes less available online. It is filtered out of the Explore feature on Instagram and is featured less prominently in Feeds and Stories on Facebook (Facebook Fact-checking, 2021). Finally, Facebook acts against repeat offenders and pages, websites, and users who constantly share misinformation will face restrictions (Facebook Fact-checking, 2021). The European External Action Service (EEAS) notes that the

initiative includes at least 70 organizations so far with fact checkers from the Associated Press, Factcheck.org, Snopes, and ABC News now part of Facebook's project.

Flagging COVID-19 misinformation

Intergovernmental organizations are also working against misinformation. The EU Code of Practice on Disinformation is a "self-regulatory standard to fight disinformation" developed to "ensure greater transparency of platforms' policies against disinformation within the EU" and signed by companies such as Facebook, Google, Twitter, Mozilla, Microsoft, and TikTok. In 2015, the European Council's East Stratcom Task Force was created in the EEAS. In 2016, the Joint Framework on countering hybrid threats was adopted, followed by the Joint Communication on increasing resilience and bolstering capabilities to address hybrid threats in 2018. The Hybrid Fusion Cell was created in the EEAS in 2016 as a focus for the analysis of hybrid threats for EU institutions, and in 2017 the European Centre of Excellence for Countering Hybrid Threats was established in Helsinki (Hagelstam, 2018). In 2018, the European Commission implemented measures to secure European elections. In late 2018, the EU presented an Action Plan against Disinformation, which included the creation of a Rapid Alert System and monitoring of the implementation of the Code of Practice signed by the online platforms (Hagelstam, 2018).

NATO also has responded to misinformation. The Allied Heads of State and Government said in the 2018 Brussels Summit Declaration and in their London Declaration that NATO should focus on countering misinformation. NATO works with its allies and the European Union, the United Nations, the G7 and civil society on the issue of misinformation, aiming to counter it with evidence-based public communications. It publicly refutes false claims, debunks misinformation targeting NATO, and coordinates with its partners to share information and best practices.



In Latin America, the Atlantic Council's Digital Forensic Research Lab and the Adrienne Arsht Latin America Center work with think tanks, media, and fact-checking organizations to identify and explain misinformation. The effort supports democratic institutions by promoting dialogue, research and analysis, joint action, capacity building, and digital resilience (Atlantic Council, 2020). They run several campaigns around misinformation in the region. For example, during their #ElectionWatch campaign, they worked to detect misinformation in the lead-up to elections in Colombia, Mexico, and Brazil in 2018 and 2019. The organization pointed out that:

In Brazil, Atlantic Council research conducted in real time found that disinformation comprised primarily organic disinformation—driven by polarization and a lack of trust in institutions. In Colombia, the Atlantic Council observed a similar trend, exacerbated at times by political leaders and the media's purposeful or accidental spread of false information. In Mexico, the Council found automation and artificial amplification to be more prominent.

In another example, #AlertaVenezuela is working towards a deeper understanding of the complexities of the information environment in both Venezuela and the larger region. By exposing and explaining misinformation in this context, the Atlantic Council, in partnership with civil society and independent media organizations working on Venezuela, will inform approaches to combatting misinformation and external interference around South America's largest modern humanitarian crisis.

Governments are also creating fact checking toolkits. In a Canadian example, a rapid response project called the COVID-19 Misinformation Portal released several resources for combatting misinformation related to COVID-19. These include Misinfo Watch, which maintains three dashboards about the pandemic, and the Botswatch "data visualization tool designed to help users, journalists, and researchers to monitor the health of coronavirus discussions on Twitter." The tool

analyzes public Twitter posts in order to flag bots that are spreading information about COVID-19. As another part of the portal, Ryerson University's Social Media Lab is leading the mapping of fact checking activities worldwide in partnership with the World Health Organization (WHO). This COVID-19 Fact Checkers Dataset "is a comprehensive international repository of over 200 active fact-checking groups and organizations that verify COVID-19 misinformation."

Many of these verification tools to establish facts proliferated during the COVID-19 pandemic. In this context, the World Health Organization (WHO) developed a mythbusters page so that readers can verify a rumour about COVID-19 that they may be unsure about. They also put out a seven-step process for people to be able to gauge misinformation they read online. WHO, partnering with Viber, created an interactive chatbot used to send accurate information directly from WHO about COVID-19 to users. In another manner, in partnership with WhatsApp and Facebook, WHO created new messaging services to link people with information about COVID-19 immediately from WHO. WHO's work on misinformation goes beyond these three companies. They state that:

Now, WHO is working with more than 50 digital companies and social media platforms including TikTok, Google, Viber, WhatsApp, and YouTube to ensure that science-based health messages from the organization or other official sources appear first when people search for information related to COVID-19. Even the dating app Tinder now features WHO health reminders, because social distancing is still appropriate during a date.



The WHO has also put together a resource list showing people how to report rumours to key social media platforms. The United Nations has produced its own anti-misinformation initiative related to COVID-19 misinformation, which is called Verified, in order “to encourage us all to check the advice we share.” The campaign’s goal is to limit misinformation by encouraging people to pause before sharing something online. As their website states, “the simple act of pausing before you share interrupts our emotional response, it triggers a moment of critical thinking.” As a part of the Verified initiative, the United Nations launched the #PledgeToPause and #TakeCareBeforeYouShare campaign. UNESCO also compiled its own resources to combat misinformation.

Content verification, computational approaches, and AI-based initiatives

A 2017 study from Yale University observed that fact checking and tagging inaccurate news stories on social media has only a small impact on whether readers believe headlines. In some cases, it could actually be counterproductive since flagging misinformation might ultimately lead users to believe other questionable content because they will assume that anything which has not been flagged must be true. In order to overcome this limitation, many initiatives go beyond fact checking. These include, among others, computational approaches, AI-based initiatives, and education campaigns. For instance, authenticating or securing media provenance is an important tool for countering misinformation. An initiative called Project Origin was started by the BBC, CBC/Radio-Canada, The New York Times, and Microsoft. As the Project Origin website notes, “positive authentication of the provenance of legitimate news stories will help by making it easier to identify manipulated and synthetic content.” Project Origin is meant to be a “new media

provenance tracking process, aimed initially at news and information content. At scale, this process could encompass traditional publishing (electronic and print), information technology, social media, and consumer software.” Their intention is to work with text, video, images, and audio content to build a global standard for establishing content integrity. The group endeavours to create a system that allows for digitally signed links that provide verifiable tracing of media content back to its publisher. Project Origin states that:

The intention of the Origin approach is to establish a chain of provenance from the point of publishing to the point of presentation. We intend to accomplish this via cryptographically-secure signatures and hashes where these are preserved in the metadata of transcoded files. Where it is not, our approach could leverage fingerprinting and watermarking techniques—or a combination of both approaches. Media data and their cryptographic hashes can be registered on a ledger, which is tamper-proof and secured by distributed ledger technology.

The authors state that the distinguishing feature of Project Origin is that it will be an industry-wide initiative that goes beyond the capabilities of what one social media company could accomplish on its own.

In another AI related example, Shu et al. (2020) have mapped out recent misinformation detection techniques developed by researchers who use AI to develop threat models and combat misinformation. By learning how users react to misinformation in comparison to true information, researchers can develop early detection tools that get ahead of widespread misinformation campaigns.

In 2020, Woolley wrote an article for the MIT Technology Review reflecting upon how AI can play a role in combatting AI-generated misinformation. He gave the example of Jigsaw, a company which designed and



built an AI-based tool called Perspective to combat online trolling and hate speech by enabling developers to automatically detect it. Woolley also highlighted the similar Deeptext, which Facebook launched in 2016. In another example, Logically is a company that counters misinformation with AI-based technology. Their four-step approach is:

- Analyze – Using advanced AI algorithms to develop systems capable of detecting and analyzing harmful content
- Alert – Responsive technology notifies users about problematic content, helping them to deploy countermeasures and mitigate the risks that it poses
- Verify – Technology works alongside expert analysts and a large, dedicated fact-checking team to shed light on dubious claims
- Mitigate – Technology and workflows detect problematic content and recommend and deploy countermeasures such as fact-checking, strategic communications, and takedowns

Logically offers its approach through an app and internet browser extension. Other services they offer include: Fact Checking as a Service (FCAAS); Fighting Extremist Content; and Intelligence Reporting.

However, Woolley concluded that AI cannot get the job done on its own and that human labour and oversight will be necessary for the foreseeable future.



Case Study 1 - Topic analysis of rumours database across all projects since 2014

One of SP's most important misinformation management tools is its WikiRumours software. WikiRumours is a web-based database which enables geographically distributed teams to collaborate asynchronously on monitoring, verifying, and countering the spread of harmful rumours and misinformation. The software is free and open-source under an MIT licence, which means that others are welcome to integrate it into their own misinformation management work. SP has deployed WikiRumours in all its misinformation projects, starting with Una Hakika in Kenya and including both Hagiga Wahid in Uganda and South Sudan, and Kijiji Cha Amani in the Democratic Republic of the Congo. It is very important to note that, while this work is technologically facilitated, the human element is still critical for it to work effectively. Members of the public report information to the project and team members with a variety of roles then add these reports to WikiRumours and proceed with the verification and counter-messaging procedures. Their roles within WikiRumours are outlined in the following table.

Role	Responsibilities
Community member	Enters and annotates rumours in the system
Proxy	Enters rumours on behalf of other users who may not have direct connectivity to the WikiRumours platform and report through other channels (e.g. SMS, voice calls, social media)
Moderator	Triages new rumours and assigns to community liaisons; can update rumour status
Community liaison	Updates rumour status
Administrator	Assigns varying levels of permission for modifying rumours, users, website content, etc.

Table 3

This process also involves gathering information from different sources and trying to make sense of it while mapping subsequent rumour reports in order to see how they develop and spread both geographically and over time. The diagram below shows how this process works on the ground.



Case Study 2 - Una Hakika in Kenya

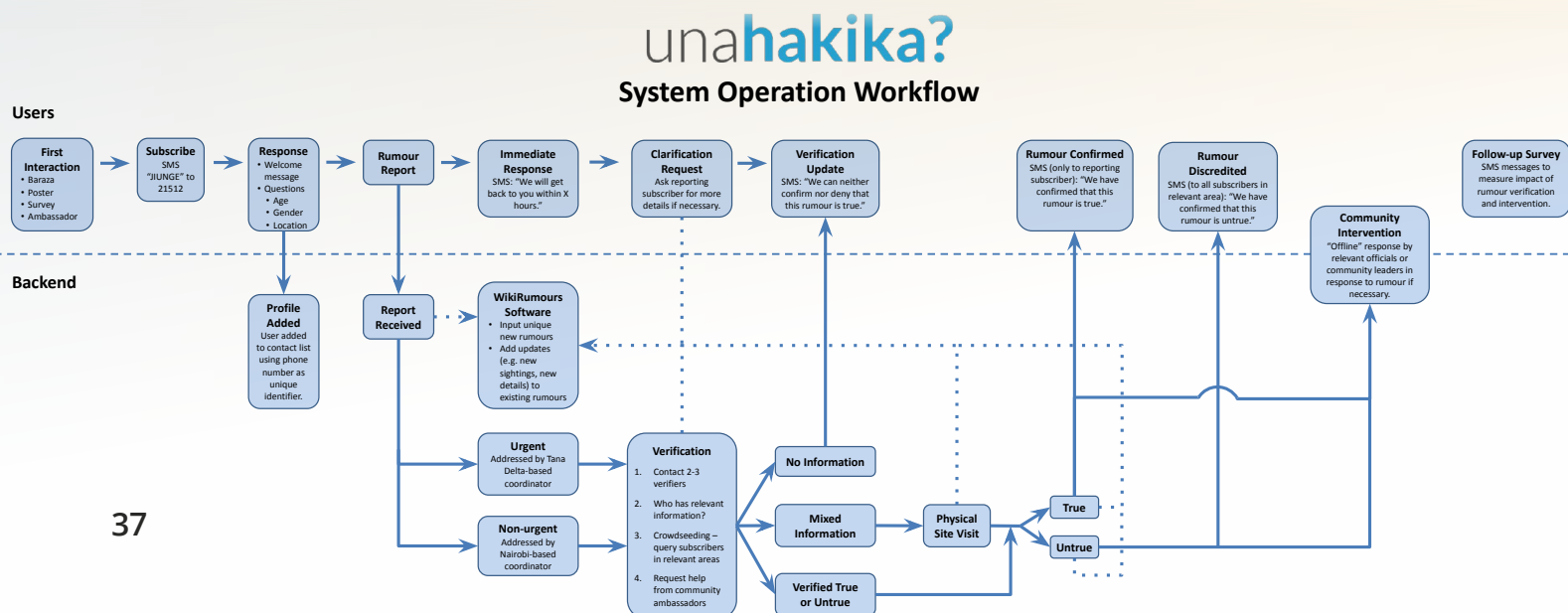
Una Hakika (UH) was SP's first misinformation management project and served as the model for all others that came after it. UH operates as a mobile phone-based information service that engages members of the public in monitoring, verifying, and countering the spread of harmful rumours and misinformation that contribute to conflict. In addition to filling information gaps in response to self-identified community needs, UH aims to encourage positive behavioural change by fostering community attitudes that prioritize critical thinking and information verification when confronted with rumours. A critical element of this work is the maintenance of a two-way flow of information with participating communities, much of which happens through SMS (short messaging service, also known as text messaging) as well as voice calls and interaction with volunteer community ambassadors. The project team then verifies and counters these rumours in accordance with the three-step process described in this report.

The development of UH was informed by an understanding of the human factors behind rumours and misinformation. Much of the theoretical foundation for Una Hakika was based upon an understanding of rumours derived from the work of Professor Nicholas

DiFonzo at the Rochester Institute of Technology. Specializing in the psychology of rumours, Professor DiFonzo's work forms the basis of three key elements of the Una Hakika theoretical framework, which are outlined below.

1. Rumours often originate in organic "sensemaking" efforts as people try to understand the world around them using incomplete information under sometimes stressful conditions, though they may also serve political, cultural, or socio-economic purposes.
2. Levelling and sharpening are twin processes whereby rumours lose details deemed to be unnecessary or nuanced while amplifying key points which resonate more with their human transmitters.
3. Group norms about acceptable standards of evidence heavily influence both the environment in which rumours and misinformation proliferate and are often connected with a lack of reliable information.

The diagram below shows the workflow that guides Una Hakika and similar projects.





Lessons learned from Una Hakika

The Una Hakika framework's working hypothesis posits that social networks—whether traditional- or technologically-based—in conflict (and other crisis) zones amplify misinformation, thus contributing to the risk of violence and hindering response and development efforts. There are three lessons related to this hypothesis.

First, this work has established the value of establishing misinformation management systems for mitigating the harmful impact of rampant rumours, whether in the form of misinformation or deliberate misinformation. Citizens, NGOs, government bodies, and the media can benefit from the clarified information environment created by such systems; they should be widely integrated into development efforts and government programming.

Second, information and communications technologies develop competencies that enable misinformation management systems to operate more efficiently than otherwise possible. The first steps in this direction were taken with the creation of the Una Hakika SMS reporting service and the implementation of the WikiRumours software that streamlined workflows so that a small number of staff could conduct work that would generally require a much larger team.

Third, systems such as Una Hakika can be made sustainable in several ways ranging from encouraging attitudinal and behavioural shifts to setting up self-sustaining, community-funded mechanisms. At the most basic level, the Una Hakika model can impart lasting changes in how communities address unverified information and knowledge of the damaging effects of misinformation on community security, personal safety, and economic stability.

Una Hakika illustrates that by refining misinformation management tools and techniques, with a focus on scalability and sustainability, this project will ensure that other actors can replicate Una Hakika for their own needs. The knowledge and the tools developed using

Una Hakika can be employed by organizations involved in peacebuilding efforts in other countries worldwide. The potential benefits of Una Hakika scaling up include impacts on how governments respond to violence prompted by misinformation. It will also assist emergency response and humanitarian missions in addressing pernicious rumours arising in chaotic settings, and create open channels between development agencies and their beneficiaries to facilitate more effective communication at all times but especially during crises. Una Hakika has already been replicated for SP's other projects such as Hagiga Wahid, A Peaceful Truth, and Kijiji Cha Amani.

Case Study 3 - Hagiga Wahid and Uganda and South Sudan

Hagiga Wahid, which is based on the Una Hakika model, provides information verification and misinformation management services to South Sudanese refugees and host communities in Rhino Camp Refugee Settlement in northern Uganda and in South Sudan. In November 2018, SP interviewed key informants, held meetings with local stakeholders, conducted focus group discussions, community meetings, and executed a baseline survey before implementing this project initially in Rhino Camp. The overarching goal was to understand refugee challenges and concerns in Rhino Camp, and understand how people there access, think about, and share information as well as how they are affected by rumours and misinformation. This preliminary work revealed great potential for a project like Hagiga Wahid to improve access to information and thus contribute to peace, stability, and development for both refugees and host communities.

The baseline survey was completed by people living in several villages in Rhino Camp and neighbouring communities in northwestern Uganda. The survey results provided valuable insights on the information needs and challenges of both South Sudanese refugees and Ugandan host community members and how those factors affected intercommunal tensions and their



relationship to the conflict in South Sudan. It particularly showed that refugees in Rhino Camp generally lack information and have little or no access to news, especially regarding local-level issues. This finding is particularly alarming because 73.6% of respondents reported believing that rumours strongly contribute to violence and instability. Several other themes were also apparent in the survey results.

- **Trust and accessibility** – Radio is the most trusted source of information among respondents, though relatively few people own radios compared to mobile phones, which 61.9% of people reported owning. This tendency, together with the general preference for communicating using voice calls, as opposed to SMS and other text-based means of communication, may be related to low literacy levels. Regardless of the cause, this set of preferences had clear implications for a mobile phone-based project such as Hagiga Wahid.
- **Rumour prevalence** – The survey results show rumours as being very common in the camps, with 58.4% of respondents reporting that they had heard a rumour over the previous 12 months. The majority of respondents (58.5%) said they had not checked to see whether such rumours were true.
- **Information confidence** – Most respondents felt either well informed or moderately informed about events within their villages or clusters since they live in small localized communities. However, the level of information confidence dropped off significantly when moving beyond the village level, with approximately 10% or less saying that they felt well -informed about the zone/district level, areas outside of the camps, and South Sudan. Most notably, nearly half of all respondents reported being uninformed about the situation outside of the refugee camps. More than half said that they were not at all informed about the situation in South Sudan.
- **Ethnicity-based trust** – An interesting aspect explained by some respondents is that even

when people do not know whether a rumour is valid, they tend to take them seriously and often believe them to be accurate. This belief in the truth of rumours may sometimes be influenced by the nature of the person from whom they are hearing the rumours, even if the ultimate source is unknown. Ethnicity plays a large role, with 43.7% of respondents highly trusting information received from members of their own ethnic group. Only 3.6% of respondents reported highly trusting information received from members of other ethnic groups and 35.6% of respondents do not trust it at all.

Case Study 4 - Kijiji Cha Amani in the Democratic Republic of the Congo

The Democratic Republic of the Congo (DRC) has a long history of political instability and violence, especially in the eastern parts of the country, where numerous armed groups constantly fight against both the government's armed forces and each other while often terrorizing the civilian population. At the same time, the country has experienced other severe crises, such as outbreaks of the Ebola virus, which have not only killed many people and threatened overall public health but have also induced a general atmosphere of stress and uncertainty. This situation has combined with the generally information-deprived nature of many parts of eastern DRC to produce conditions which are ideal for rumours and misinformation to arise and spread. Such rumours then make it difficult for citizens to make effective decisions about how to live their lives and navigate sometimes dangerous situations. There is also evidence that they contribute to the tensions which can lead to intercommunal conflict and may also sometimes relate to instances of mob violence against suspected criminals.

These conditions led a local community leader to



approach the Sentinel Project after learning about the successful Una Hakika project in Kenya with the goal of replicating it in eastern DRC. This collaboration resulted in the Kijiji Cha Amani (KCA) project, which is Swahili for “Peace Village,” and now engages more than 12,000 subscribers in the process of monitoring, verifying, and countering the spread of harmful rumours and misinformation. While the project is still in a relatively early stage of its implementation, Kijiji Cha Amani (KCA) has demonstrated that there is significant demand for participatory misinformation management mechanisms in DRC. It is also noteworthy that KCA has played a localized early warning and public safety alert role in cases of ongoing violence, which are common in the project area. The SP team now plans to significantly scale up KCA in order to increase its reach and impact.

Discussions and conclusions

In response to the first research question, the background research and case studies reveal three central considerations. First, it is clear that human input into the misinformation management process continues to be important for the success of such efforts, which will realistically continue to be labour intensive for the foreseeable future, with automation only reducing the workload of human moderators and verifiers but not replacing them in any meaningful way.

Second, initiatives must be culturally relevant and the lessons shared in the 2016 Una Hakika evaluation report remain valid in this respect. Misinformation management systems cannot be only imposed from above. Instead, they must be implemented by entering into communities using culturally relevant introduction processes followed by cooperative efforts. Misinformation management efforts should not replace existing communication practices in a new environment but rather adapt to the variety of both low-technology and high-technology methods that are already used in the identified area.

Third, trust is one of the most important but also most difficult components to establish and maintain during a misinformation management project. A project which lacks transparency will promptly lose the majority of its value to the community and to the pursuit of peace. Suspicion or hostility from some community members is a natural occurrence but trust can generally be built and maintained with the majority if misinformation management projects maintain good relations.

These considerations make it difficult to scale or universalize projects that require intensive on-the-ground knowledge of the local context. They point to the enduring relevance of low-technology interventions, irrespective of large-scale joint initiatives by inter-governmental organizations, universities, public institutions, technology-heavy content verification, computational approaches, and AI-based initiatives. Indeed, the examples of SP's projects demonstrate the intricate social dynamics which influence misinformation management efforts, the highly contextual nature of how these approaches can be implemented, and the importance of building user trust. Any attempts at scaling up or reconciling micro and macro approaches should take these points into consideration. Some large organizations recognize this point. For instance, UNESCO has partnered with radio stations to counter misinformation related to COVID-19 but producing audio messages that can be used by radio stations worldwide and making them freely available to help stop the spread of false information. In a country-specific example, the WHO provided equipment and training to journalists, producers, and directors of health programs at radio stations in Côte d'Ivoire to deliver accurate information to help stop the spread of the virus.

မြမ်းချမ်းသော အမှန်တရား သည်
ရေစိုညှိမြန်းချမ်းသာယာမှုအတွက်
မှန်ကန်တိကျသောသတင်းအချက်အလက်မျိုးကို
ပေးအပ်ခြင်းဖြင့် ထောက်ပံ့သည်။
မြမ်းချမ်းရေးကို အမှန်တရားဖြင့်
လူထုစိမ့်ချက်ဖြစ်သည်။
တည်ဆောက်နိုင်သည်။ အမှန်တရားကိုလည်း
သိကျမှန်ကန်သည်။ သတင်းအချက်အလက်ဖြင့်
တင်ပြဆောင်သွားနိုင်သည်။
ပါဝင်ရန်

SMS to 77770701

အမှန်တရား
Truth



Section 3: Main Findings – Part 2

Who is responsible for leading misinformation management efforts? Misinformation is a phenomenon with vast reach, making it the responsibility of all facets of society. The United Nations published a list of recommendations for addressing hate speech related to COVID-19.

Their recommendations extended to other UN agencies, governments, technology companies, media, civil society organizations, and other actors. The Atlantic Council's report on Latin America outlined how "government, technology and social media companies, telecommunication companies, fact-checking organizations and the media, civil-society organizations and academic institutions, and international institutions" are responsible for countering and mitigating misinformation (p. 42). This section answers the question, "How can social media platforms and related government policies reduce the ease of proliferation for misinformation and the degree to which it has negative impacts?"

Technology companies

It is increasingly evident that misinformation will increase in prominence and prevalence, including as a contributing factor for violence and instability. Any potential solutions that aim to have a large-scale impact need to incorporate the parties that operate the means of dissemination—namely, social media companies. What can these companies do? What are the best means for influencing them to do so? Corporations that operate social media platforms usually have more legal freedom to engage in surveillance and censorship. Still, they are often reluctant to do so for a variety of reasons. Even when inclined to address issues like hate speech and misinformation, such companies struggle

to find an appropriate global standard since their users typically span numerous legal and regulatory jurisdictions as well as cultural contexts, not all of which align easily with the Western liberal democratic culture in which companies like Facebook and Twitter were created and typically develop their policies. As reported by The Guardian, Paul Chichester, Director of Operations of the UK's National Cyber Security Centre, has stressed the need for social media companies to do more to combat misinformation.

For technology companies, promising avenues continue to include flagging misinformation content and developing programs to identify and remove cyber troops and bot accounts. They can also take decisive steps and develop policies regarding misinformation and hate speech that are found on their platforms. For example, in late 2020, the Washington Post reported that Facebook had announced its intentions to "remove false claims that could include misinformation about the safety, efficacy, ingredients or side effects of the [COVID-19] vaccines." The New York Times reported that Facebook also prohibited political advertising in the lead-up to the 2020 US presidential election. In November 2020, Facebook banned a "Stop the Steal" group on its platform due to growing signs that members would incite violence. In December 2020, Twitter expanded its hateful conduct policy. On the company blog, Twitter Safety noted that:



While we encourage people to express themselves freely on Twitter, abuse, harassment and hateful conduct continue to have no place on our service. In July 2019, we expanded our rules against hateful conduct to include language that dehumanizes others on the basis of religion or caste. In March 2020, we expanded the rule to include language that dehumanizes on the basis of age, disability, or disease. Today, we are further expanding our hateful conduct policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin.

Indeed, the company blocked British conspiracy theorist David Icke in November 2020 and permanently blocked former US president Donald Trump in January 2021. As the 2020 Cyber Troops Report by the Oxford Internet Institute notes, “Public announcements by Facebook and Twitter between January 2019 and December 2020 reveal that more than 317,000 accounts and pages have been removed by the platforms.” In another example, due to the COVID-19 pandemic, WhatsApp instituted a new rule that does not allow heavily forwarded pieces of information (i.e. messages that have already been forwarded five times) to be further forwarded to more than one individual at a time. These initiatives highlight the potential of policy approaches by technology companies but it is important to also recognize their limitations when it comes to implementation, which may be supported by technological tools but still requires significant human input.

Government interventions

There are some very compelling reasons why governments should aim to reduce misinformation, especially when there are clear links between misinformation and physical violence. Government policies are essential since some degree of legislation or regulation is generally necessary to compel private sector action. Government policy interventions therefore remain a core element of potentially reducing the proliferation of incendiary rumours and misinformation, which increasingly propagate on social media platforms controlled by

the private sector. Poynter keeps track of government regulations regarding misinformation and they have compiled a list of the actions taken by 98 countries, which are also categorized by the type of action, each country's focus, and its orientation towards approaches such as monitoring, awareness, and sanctions.

The Library of Congress legal website has published a report on the actions that 15 different governments have taken against misinformation. As the author Tariq Ahmed (2019) notes, in Canada, the right to freedom of expression is protected by subsection 2(b) of the Canadian Charter of Rights and Freedoms. Everyone has the fundamental freedom of “thought, belief, opinion and expression, including freedom of the press and the media communication. The prohibition of the distribution of false news would only be against the law if “it is defamatory and covered by libel laws.” As he notes, Section 181 of the Criminal Code prohibits the spreading of false news. However, the Supreme Court of Canada held in the case of *R. v. Zundel* (1992) that the law was unconstitutional because it violates freedom of expression (Ahmed, 2019). As Ahmed (2019) continues, in order to manage this unresolved quandary, there are government laws, policies, and initiatives that toe the line of avoiding overregulation that curtails freedom of speech while still reducing the most egregious types of misinformation that cause the destabilization of democracy and the proliferation of violence. Similarly, as one report from Johns Hopkins University states, “governments in free societies should work within the parameters of free speech and expression to build resiliency and create deterrents” (AICGS, p. 39).

In 2019, Canada introduced a Digital Charter that outlines ten principles to protect Canadians online. Regarding misinformation, the eighth principle promises that the government of Canada “will defend freedom of expression and protect against online threats and disinformation designed to undermine the integrity of elections and democratic institutions.” The ninth principle states that “Canadians can expect that digital platforms will not foster or disseminate hate, violent extremism or criminal content.” In another initiative, in 2019 Canada



signed onto the Christchurch Call to Action in which governments and internet service providers pledged to eliminate terrorist and violent extremist content online. The commitments include:

...building more inclusive, resilient communities to counter violent radicalization; enforcing laws that stop the production and dissemination of terrorist and extremist content online; and encouraging media to apply ethical rules when reporting on terrorist events to avoid amplifying terrorist and violent extremist content.

The following sections explore the actions that governments can take to address misinformation in the realms of protecting elections and encouraging media literacy.

Elections

There is a moral imperative to protect democratic ideals such as free and fair elections, which is why the Canadian government established the Canadian Security and Intelligence Threats to Elections Task Force in order to prevent “covert, clandestine, and criminal activities from interfering with or influencing the electoral process.” It comprises members of the Canadian Security Intelligence Service, the Royal Canadian Mounted Police, and Global Affairs Canada. The government also released Canada’s Declaration on Electoral Integrity Online with the intention “to ensure integrity, transparency, and authenticity, subject to Canadian laws and consistent with other legal obligations” in the lead-up to the 2019 federal election. In May 2019, Global News reported that several major technology companies, including Facebook, Google, and Microsoft, had signed onto the Declaration.

In another example, the Elections Modernization Act included a point on political advertisements on social media platforms. The backgrounder to the new act outlined how:

Canada became one of the first countries to require major online platforms to maintain a

registry of partisan and election advertising published during the pre-election and election periods. The registry must include a copy of the advertising message, and the name of the person who authorized it. This complements the obligation on political parties and third parties to identify themselves on their partisan and election advertising during these periods.

CTV News outlined that the new act also has a provision that “makes it an offence to make false statements about a candidate to influence the outcome of an election.” Although this provision is somewhat unclear, there is some potential benefit in regulating some aspects of misinformation around elections in order to ensure that they are protected from interference and manipulation, though this must be done with as much restraint as possible.

Media literacy

Shu et al. (2020) also pointed out that there have been many initiatives aimed at educating people about misinformation in order to reduce its dissemination. They give the example of the News Literacy Project in the United States, which is:

...a nonpartisan national education nonprofit, [that] provides programs and resources for educators and the public to teach, learn and share the abilities needed to be smart, active consumers of news and information and equal and engaged participants in a democracy.

Shu et al. (2020) also gave the example of Finland, which has educational programs teaching students how to evaluate the authenticity of online articles. The Guardian has described how Finnish schools have started teaching students about misinformation in primary grades. The article details how:

The curriculum is part of a unique, broad strategy devised by the Finnish government after 2014, when the country was first targeted with fake



news stories by its Russian neighbour, and the government realised it had moved into the post-fact age.

According to Shu et al. (2020), further improving analytical and actively open-minded thinking might reduce people's consumption of misinformation because critical thinking involves contemplating alternative explanations and drawing on compelling evidence (Bronstein et al., 2019; Pennycook & Rand, 2019a, 2019b in Shu et al., 2020). Some initiatives are therefore geared towards building critical thinking skills. The Government of Canada's Digital Citizen Initiative is another strategy to build citizen resilience against online misinformation. In 2019, the Canadian government announced plans to "build citizens' critical thinking and preparedness against online misinformation, and other online harms." The aim is to encourage "civic, news, and digital media literacy through third-party educational activities and programming to help citizens become resilient against online harms." Its goal is to use various programs to help citizens learn to:

...critically assess online information; understand how algorithms work and when they might impact a user's online experience; recognize how and when malicious actors exploit online platforms; acquire skills to avoid being susceptible to online manipulation, and effectively engage in public debate and online discussions.

Over 20 projects have been funded to implement related programming through this initiative, though it is likely too soon to evaluate whether or not they have been successful in achieving their goals.

Civil society

Civil society can play a very significant role in countering misinformation. Beyond influencing private sector technology policy, advocating for government intervention, and providing media literacy education, civil society actors also engage in a variety of other activities. For example, the Johns Hopkins report states that civil society can:

[First] ... act as a watchdog, policing social media and exposing disinformation campaigns as they emerge. Second, it can help to inoculate publics against information manipulation by supporting education outreach and media literacy programs. Third, it can apply pressure to tech companies, businesses, and advertisers that wittingly or unwittingly host, support, or incentivize creators of false and misleading content. Finally, civil society can work with governments, the media, and each other to improve the conditions of mistrust and polarization that create fertile breeding grounds for the spread of disinformation.

Freedom House notes that:

Besides the stated intentions of social networks themselves, several publications and other pieces of research have argued that social networks must cooperate with civil society in order to effectively fight against inauthentic or other harmful behavior. In addition, the EU Code of Conduct on Countering Illegal Hate Speech Online, a document that is not legally binding but which has been signed by the biggest internet intermediaries operating in Europe, encourages partnerships between tech companies and civil society organizations in the fight against illegal content online.

The Kennedy School Review cites examples of civil society efforts to address misinformation. The authors give the example of the platform CoFacts in Taiwan, which has created a chatbot that uses crowdsourcing to fact check messages on the mobile messaging application called Line. They note that:

When Line users encounter a potential piece of misinformation, they can forward the message to the CoFacts chatbot. The chatbot returns a result based on a database of previously forwarded false messages. But if no match exists, a volunteer editor reviews the message content. The editor then assigns a rating and communicates the finding to the Line user. Editors flag false stories



with one of four possible ratings: not applicable (not related to fact-checking), personal opinion (not verifiable information), correct message, or false message.

The authors further state that this model is flexible and scalable, which are important characteristics for any misinformation management system to be truly impactful.

Humanitarian organizations

It is also important to highlight the ways in which humanitarian actors such as SP can also play a role in misinformation management. For example, the Accountability Lab was founded in 2012 to work with young people on developing new ideas for promoting accountability, transparency, and open government. Its global networks of local Accountability Labs “shift societal norms, solve intractable challenges, and build ‘unlikely networks’ for change.” Regarding misinformation, the Accountability Lab’s Coronavirus CivActs Campaign in Nepal gathers rumours, concerns, and questions from communities “to eliminate information gaps between the government, media, NGOs, and citizens. By providing the public with facts, the CCC ensures a better understanding of needs regarding the coronavirus and debunks rumours before they can do more harm.” These are just a couple of the many examples that demonstrate the potential role of humanitarian organizations in this sector.

Crowdsourcing

Dr David Rand of the MIT Sloan School of Management has stated that “Crowdsourcing, if implemented correctly, is a promising approach to fighting the spread of misinformation and false news.” He was referring to a study that he conducted with Dr Gordon Pennycook in 2019, which found that “crowdsourced trust ratings can effectively differentiate more versus less reliable sources.” In the press release related to the study,

Rand explains that “Our study is good news because we find a scalable solution to this problem, based on the surprisingly good judgment of everyday Americans. Things may not be as hopeless as most coverage of fake news makes you think.” While SP has extensive experience with crowdsourcing rumour reports from members of the public in various countries, this MIT study highlights the interesting potential of engaging the general public directly in the information verification process itself, which may offer another significant tool for addressing misinformation.



14/05/2021

DAD. /20pts
Q. Sur format A₄, au sens
Vertical, échelle 1/1, de
sirez une fleur et indi-
quez-y toutes ses parti-



La
de
Ens
a sta
Contre
des f
trans
préc
www
+212



Section 5: Recommendations

This report ends on a cautionary note because misinformation is a complex problem for which there is no single solution that can be applied in all contexts.

Regarding technology company policies, the private sector must continue to address misinformation issues, both because it is a moral imperative to mitigate misinformation's most dangerous consequences and because there is arguably a business case to be made for creating safer platforms and healthier online communities. There are many tools available to such companies, such as further developing fact checking and provenance verification tools. Although the effectiveness of such tools is debatable, it remains essential to flag and remove categorically false information. While fact checking may seem like a rabbit-chasing exercise, it is better than not having fact checking at all, regardless of the currently inconclusive research on its effects. Technology companies should particularly continue to crack down on bot-based misinformation, which can be created and disseminated at speeds far exceeding what human users can achieve, thus having potentially far greater negative impacts.

Another vital facet of misinformation management is to pay attention to non-English content, especially in countries experiencing political turbulence and violence. Most social media platforms are relatively adept at monitoring and moderating content in English and other major languages but this focus can overlook important content that is published in less widely-spoken languages. Of course, this depends on context, but the most egregious forms of misinformation that lead to violence on a mass scale, such as in Myanmar, should receive the most attention from mitigation efforts despite being disseminated in languages with which technology companies are generally unfamiliar. Addressing such situations requires greater investment not only in technology but also in human resources. It

is important to note that there is also a growing need to pay attention to non-text forms of misinformation, such as fabricated and manipulated images and videos, which threaten to undermine trust in visual content while also potentially evoking stronger emotional responses in viewers when compared to what is seen in response to text-based content.

Governmental misinformation management efforts require striking a balance between legislation and regulation on the one hand and fundamental rights and freedoms on the other. This report has predominantly examined the promising governmental initiatives currently underway in Canada because policy recommendations are difficult to develop for several key reasons. First, any universal policy recommendation is problematic due to the high level of variation between jurisdictions in terms of cultural norms, political conditions, and societal norms such as respect for institutions and civil liberties. Second, much like when dealing with hate speech, managing misinformation through government policy measures introduces a tension between the desire to reduce the prevalence of the dangerous phenomenon and the need to preserve freedom of expression. While many government efforts to counter misinformation may be sincerely intended to serve the public interest, this is not a problem that can be easily controlled. Any government action should be weighed against the potential loss of digital freedom that it introduces.

While some observers will view any concerns about freedom of expression to be thinly veiled attempts to defend the disseminators of harmful content, there are in fact genuine questions that need to be asked



about whether or not governments should be playing a central role in such efforts. This question is relevant everywhere in the world, including liberal democracies, but becomes even more salient in countries that lack strong protections for freedom of expression. There have already been numerous examples worldwide of governments developing strict legal penalties for disseminating so-called “fake news,” which is often used to stifle dissent and target journalists who report on government misconduct. It is therefore important to exercise caution and restrain government action since overreactions to misinformation can counteract and even undermine democratic ideals.

Some observers may react strongly when they hear arguments for government restraint that are based on freedom of expression. As mentioned above, such arguments may seem like defences of extremism by people who care more about freedom of expression and freedom of speech than preventing violence. However, misinformation management is not a clear-cut exercise which is naturally led by benign actors. Except for stopping explicit incitement of violence, the least possible amount of government intervention is recommended in order to avoid infringing upon fundamental human rights. Even if the authority to regulate misinformation is not exploited, the government and courts should not be the sole arbiters of “truth,” a situation which would present a potentially slippery slope from which it would be very difficult to recover. Furthermore, single-minded thinking about the government’s role might be counterproductive since governments getting heavily involved in misinformation may reinforce conspiracy theory types of thinking among those who are prone to believe misinformation. Government moves to censor misinformation may in fact lead some people to believe even more strongly in the truth of the content in question. Moreover, governments should not be considered to be neutral parties since even liberal democracies generally act in line with their own interests. While some of these efforts may be well-intentioned attempts to counter harmful misinformation, others are almost certainly moves by authoritarian regimes to stifle dissent and

control public discourse.

As noted in this report, governments often spread misinformation in order to advance their own strategic interests. For example, Freedom House studies digital authoritarianism and assesses the degrees of digital freedom found in various countries, and found that digital freedom had generally declined during their most recent reporting period. The authors of Freedom House’s most recent study note that “of the 65 countries assessed, 26 have been on an overall decline since June 2017, compared with 19 that registered net improvements.” In particular, in 2020, their Freedom on the Net report observed that the COVID-19 pandemic had expedited digital repression. The authors write that “For the 10th consecutive year, users have experienced an overall deterioration in their rights, and the phenomenon is contributing to a broader crisis for democracy worldwide.” Although liberal democracies may not engage in misinformation dissemination based on their own self-serving interests as often or as explicitly as authoritarian states, all governments will generally act in ways which serve their interests for self-preservation to some degree.

It is important to emphasize that the cautionary points outlined above do not mean that governments should not do anything to counter misinformation. Indeed, many governments are already taking decisive and positive steps and, as this report has shown, there remain many opportunities to manage misinformation that do not involve legislative action. For example, SP recommends that governments should continue to fund media literacy campaigns such as the Digital Citizen Initiative. While governmental authorities can highlight the harms of online misinformation, citizens can and should generally be left to make their own decisions about what to believe on the internet. Governments can also require transparency from social media companies regarding their handling of hate speech and the incitement of violence, both of which can overlap with misinformation. When there is a clear link from a type of misinformation circulating to hate speech and physical violence, social media companies should have



protocols in place to swiftly and transparently address such content.

Due to misinformation's multifaceted nature, there is no universal remedy that will solve this growing problem. Thus, after concluding the process of developing this report, the SP team ultimately supports the recommendations in the Kennedy School Review article about focusing on supporting civil society actors in their own efforts to address misinformation. However, despite the importance of increasing support for civil society in this regard, it is also true that no one sector will be able to address misinformation by itself. As the authors of that article note, "Rather than searching for

a sweeping top-down solution, an effective strategy may involve identifying and supporting a diverse field of organizations that are committed to addressing this issue." The authors of this report certainly agree that such a pluralistic and collaborative approach is likely to be the most effective.





Works cited

- A guide to anti-misinformation actions around the world. (2019, August 14). Retrieved February 16, 2021, from <https://www.poynter.org/ifcn/anti-misinformation-actions>
- Adams, T. (2018, November 18). The charge of the chatbots: How do you tell who's human online? Retrieved March 22, 2021, from <https://www.theguardian.com/technology/2018/nov/18/how-can-you-tell-who-is-human-online-chatbots>
- Ahmad, T. (2019, September 01). Canada. Retrieved February 16, 2021, from <https://www.loc.gov/law/help/social-media-disinformation/canada.php>
- Al Jazeera. (2017, August 05). How fake news could shape Kenya's elections. Retrieved February 16, 2021, from <https://www.aljazeera.com/program/the-listening-post/2017/8/5/how-fake-news-could-shape-kenyas-elections/>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. doi:10.3386/w23089
- An independent assessment of the human rights impact of Facebook in Myanmar. (2020, August 25). Retrieved February 16, 2021, from <https://about.fb.com/news/2018/11/myanmar-hria>
- Anand, G., & Raj, S. (2017, May 25). Rumors on WhatsApp Ignite 2 Mob Attacks in India, Killing 7. Retrieved February 16, 2021, from <https://www.nytimes.com/2017/05/25/world/asia/india-vigilante-mob-violence.html>
- Anderson, J., & Rainie, L. (2020, August 17). The future of truth and misinformation online. Retrieved February 16, 2021, from <https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online>
- Audio resources to counter disinformation. (2020, May 05). Retrieved February 16, 2021, from <https://en.unesco.org/covid19/communicationinformationresponse/audioresources>
- AWS. (2021). What is Artificial Intelligence? Machine Learning and Deep Learning. Retrieved February 16, 2021, from <https://aws.amazon.com/machine-learning/what-is-ai>
- Bandeira, L., Barojan, D., Peñarredonda, J., & Pérez Argüello, M. (2019). Disinformation in Democracies: Strengthening Digital Resilience in Latin America. Retrieved March 22, 2021, from <https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-democracies-strengthening-digital-resilience-latin-america>
- BBC launches huge new international anti-disinformation initiative. (2018, November 12). BBC News. Retrieved March 22, 2021, from <https://www.bbc.co.uk/mediacentre/latestnews/2018/beyond-fake-news>
- Benedictus, L. (2016). Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges. The Guardian. Retrieved from <https://www.theguardian.com/media/2016/nov/06/troll-armies-socialmedia-trump-russian>
- Betsch C. (2011). Innovations in communication: the Internet and the psychology of vaccination decisions. Euro Surveill. 16(17). <https://doi.org/10.2807/ese.16.17.19849-en>
- Boutillier, A. (2018, February 22). 'Total' information warfare a threat TO democracy: CSIS report. Retrieved February 16, 2021, from <https://www.thestar.com/news/canada/2018/02/22/csis-says-total-information-warfare-a-threat-to-democracy.html>



- Bovet, A., & Makse, H. (2019, January 02). Influence of fake news in Twitter during the 2016 US presidential election. Retrieved February 16, 2021, from <https://www.nature.com/articles/s41467-018-07761-2>
- Bradshaw, S. & Howard, P. N. (2017) Troops, Trolls and Trouble Makers: A Global Inventory of Social Media Manipulation. Computational Propaganda Project Working Paper Series, Working paper 2017,12.
- Bradshaw, S. & Howard, P. N. (2018) Challenging Truth and Trust: A Global Inventory of Social Media Manipulation. Computational Propaganda Project Working Paper Series. Working Paper 2018.1
- Bradshaw, S. & Howard, P. N. (2019). The Global Disinformation Disorder: 2019 Inventory of Social Media Manipulation. Computational Propaganda Project Working Paper Series. Working Paper 2019.3.
- Bronstein, M., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. (2018). Reduced analytic and Actively Open-minded thinking help to explain the link between belief in fake news and Delusionality, dogmatism, and religious fundamentalism. *Journal of Applied Research in Memory and Cognition*, 8(1). doi:10.1016/j.jarmac.2018.09.005
- Bryden, J. (2019, February 07). Elections Canada can't do much about fake news targeting candidates. Retrieved February 16, 2021, from <https://www.ctvnews.ca/politics/elections-canada-can-t-do-much-about-fake-news-targeting-candidates-1.4287398>
- Bryden, J. (2019, May 27). Several tech giants sign onto Canadian Declaration on Electoral integrity. Retrieved February 16, 2021, from <https://globalnews.ca/news/5323084/tech-giants-electoral-integrity>
- Canada joins Christchurch call to action to Eliminate terrorist and violent extremist content online. (n.d.). Retrieved February 16, 2021, from <https://pm.gc.ca/en/news/news-releases/2019/05/15/canada-joins-christchurch-call-action-eliminate-terrorist-and-violent>
- Canada, G. (2021, January 12). Canada's digital charter: Trust in a digital world. Retrieved February 16, 2021, from https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00108.html
- Canadian Heritage. (2019, August 21). Backgrounder – helping citizens critically assess and become resilient against harmful online disinformation. Retrieved February 16, 2021, from <https://www.canada.ca/en/canadian-heritage/news/2019/07/backgrounder--helping-citizens-critically-assess-and-become-resilient-against-harmful-online-disinformation.html>
- Chazan, G. (2017, February 15). Rise of refugee 'fake News' rattles German politics. Retrieved February 16, 2021, from <https://www.ft.com/content/11410abc-ef6e-11e6-ba01-119a44939bb6>
- Civil society: A key player in the global fight against misinformation. (2020, February 06). Retrieved February 16, 2021, from <https://ksr.hkspublications.org/2020/02/07/civil-society-a-key-player-in-the-global-fight-against-misinformation>
- Code of practice on disinformation. (2021, January 18). Retrieved February 16, 2021, from <https://ec.europa.eu/digital-single-market/en/code-practice-disinformation>
- ComProp: Industrialized Disinformation: 2020 global inventory of organized social media manipulation. (2020). Retrieved February 16, 2021, from <https://comprop.oii.ox.ac.uk/research/posts/industrialized-disinformation/#continue>
- Condliffe, J. (2020, April 02). Fake news is unbelievably cheap to produce. Retrieved February 16, 2021, from <https://www.technologyreview.com/2017/06/14/151233/fake-news-is-unbelievably-cheap>



Coronavirus: Ukraine PROTESTERS attack buses carrying China evacuees. (2020, February 21). Retrieved February 16, 2021, from <https://www.bbc.com/news/world-europe-51581805>

Côte D'Ivoire: Who fights misinformation through radio. (n.d.). Retrieved February 16, 2021, from <https://www.who.int/news-room/feature-stories/detail/c%C3%B4te-d-ivoire-who-fights-misinformation-through-radio>

Countering disinformation in Latin America. (2020, June 23). Retrieved February 16, 2021, from <https://www.atlanticcouncil.org/countering-disinformation-in-latin-america>

COVID-19 Mythbusters. (2020). Retrieved February 16, 2021, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>

COVID-19 pandemic Demonstrates multilateral Cooperation key to overcoming global challenges, president stresses as General Assembly concludes annual Debate | Meetings coverage and press releases. (2020). Retrieved February 16, 2021, from <https://www.un.org/press/en/2020/ga12273.doc.htm>

DiResta, R. (2020, September 20). The supply of Disinformation will soon be infinite. Retrieved February 16, 2021, from <https://www.theatlantic.com/ideas/archive/2020/09/future-propaganda-will-be-computer-generated/616400>

Dwoskin, E. (2020, December 03). Facebook steps up campaign to ban false information about coronavirus vaccines. Retrieved February 16, 2021, from <https://www.washingtonpost.com/technology/2020/12/03/facebook-covid-vaccine>

Elliott, V. (2020, January 15). In Congo's EBOLA zone, MISINFORMATION persists even as cases slow. Retrieved February 16, 2021, from <https://www.thenewhumanitarian.org/news-feature/2020/1/15/Ebola-Congo-social-media-whatsapp-misinformation>

Elliott, V. (2020, January 15). In Congo's EBOLA zone, MISINFORMATION persists even as cases slow. Retrieved February 16, 2021, from <https://www.thenewhumanitarian.org/news-feature/2020/1/15/Ebola-Congo-social-media-whatsapp-misinformation>

Emmott, R. (2020, March 18). Russia deploying coronavirus disinformation to sow panic in West, EU document says. Retrieved February 16, 2021, from <https://www.reuters.com/article/us-health-coronavirus-disinformation/russia-deploying-coronavirus-disinformation-to-sow-panic-in-west-eu-document-says-idUSKBN21518F>

Entous, A., Timberg, C., & Dwoskin, E. (2017, September 25). Russian operatives used Facebook ads to exploit America's racial and religious divisions. Retrieved February 16, 2021, from https://www.washingtonpost.com/business/technology/russian-operatives-used-facebook-ads-to-exploit-divisions-over-black-political-activism-and-muslims/2017/09/25/4a011242-a21b-11e7-ade1-76d061d56efa_story.html?itid=ik_inline_manual_2

Establishment, C. (2019, April 09). 2019 update on cyber threats to Canada's democratic process. Retrieved February 16, 2021, from <https://www.cse-cst.gc.ca/en/media/media-2019-04-08>

Experts fear false rumours could HARM chinese cooperation ON CORONAVIRUS. (2020, February 20). Retrieved February 16, 2021, from <https://www.theguardian.com/world/2020/feb/20/coronavirus-chinese-scientists-false-rumours-experts>

Facebook for business: Business help center. (2021). Retrieved March 22, 2021, from <https://www.facebook.com/business/help/2593586717571940?id=673052479947730>



Fake news, propaganda, and Disinformation: Learning to critically Evaluate media Sources: Infographic: Spot fake news. (n.d.). Retrieved February 16, 2021, from https://guides.library.cornell.edu/evaluate_news/infographic

Final report of the high-level expert group on fake news and online disinformation. (2018, March 21). Retrieved February 16, 2021, from <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

Four Western Balkans' Fact-checking Organisations join Facebook's platform against disinformation. (2020, August 13). Retrieved March 22, 2021, from https://eeas.europa.eu/headquarters/headquarters-homepage/84059/four-western-balkans%E2%80%99fact-checking-organisations-join-facebook%E2%80%99s-platform-against_en

Ganley, E., & Satter, R. (2017, May 05). As French campaign Ends, Macron's team says it's the victim OF 'massive' hacking attack. Retrieved February 16, 2021, from <https://www.thestar.com/news/world/2017/05/05/french-presidential-race-enters-final-day-with-violent-protests-voter-angst.html>

Gottfried, J., & Grieco, E. (2020, July 28). Younger Americans are better than older Americans at telling factual News statements from opinions. Retrieved February 16, 2021, from <https://www.pewresearch.org/fact-tank/2018/10/23/younger-americans-are-better-than-older-americans-at-telling-factual-news-statements-from-opinions/>

Government of Canada. (2017, March 09). Questions and answers - an act to amend the Criminal CODE (REMOVING UNCONSTITUTIONAL portions or PROVISIONS). Retrieved February 16, 2021, from <https://www.justice.gc.ca/eng/csj-sjc/pl/cuol-mgnl/qa-qr.html>

Government of Canada. (2019, June 17). Section 2(b) – freedom of expression. Retrieved February 16, 2021, from <https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-cddl/check/art2b.html>

Government of Canada. (2020, August 17). Retrieved February 16, 2021, from <https://www.canada.ca/en/canadian-heritage/services/online-disinformation.html#a1>

Gruzd, A., & Mai, P. (2021, February 11). COVID-19 Bots Dashboard. Retrieved February 16, 2021, from <https://covid19misinfo.org/botswatch>

Gu, L., Kropotov, V., & Yarochkin, F. (n.d.). The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public. Retrieved February 16, 2021, from https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf

Guess, A., Nagler, J., & Tucker, J. (2019, January 01). Less than you think: Prevalence and predictors of fake news dissemination on facebook. Retrieved February 16, 2021, from <https://advances.sciencemag.org/content/5/1/eaau4586>

Guess, A., Nyhan, B., & Reifler, J. (2020, March 02). Exposure to UNTRUSTWORTHY websites in the 2016 US election. Retrieved February 16, 2021, from <https://www.nature.com/articles/s41562-020-0833-x>

Guge, E. (2020, Summer). Targeted Disinformation Warfare: How and Why Foreign Efforts are Effective, and Recommendations for Impactful Government Action. Retrieved February 16, 2021, from <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1645&context=creativecomponents>

Hagelstam, A. (2018, November 23). Cooperating to counter hybrid threats. NATO Review. Retrieved March 22, 2021, from <https://www.nato.int/docu/review/articles/2018/11/23/cooperating-to-counter-hybrid-threats/index.html>



Harper, C. A., & Baguley, T. (2019). "You are fake news": Ideological (a)symmetries in perceptions of media legitimacy. doi:10.31234/osf.io/ym6t5

Hayden, S. (2019, June 20). Rumors and false information surround Ebola outbreak in DRC. Retrieved February 16, 2021, from <https://time.com/5609718/rumors-spread-ebola-drc>

Henley, J. (2020, January 29). How Finland starts its fight against fake news in primary schools. Retrieved February 16, 2021, from <https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news>

Home. (n.d.). Retrieved February 16, 2021, from <https://www.takecarebeforeyoushare.org>

Home: Misinformation-Fighting, Ai-powered News & fact checking. (n.d.). Retrieved February 16, 2021, from <https://www.logically.ai>

How do I spot misinformation? (2020, November 03). Retrieved February 16, 2021, from <https://onereach.library.utoronto.ca/faq/how-do-i-spot-misinformation>

How to report misinformation online. (n.d.). Retrieved February 16, 2021, from <https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/how-to-report-misinformation-online>

How to Spend \$800,000 on Facebook to Influence African Elections. (2019, May 27). Retrieved February 16, 2021, from <https://www.ictworks.org/facebook-influence-african-elections/#.X9bImthKjIU>

How to spot fake news. (n.d.). Retrieved February 16, 2021, from <https://www.torontopubliclibrary.ca/spotfakenews>

Information sharing & countering disinformation. (2020, September 02). Retrieved February 16, 2021, from <https://en.unesco.org/covid19/communicationinformationresponse>

Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges. (6 November 2016). Retrieved February 16, 2021, from <https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian>

Isaac, M. (2020, September 03). Facebook moves to limit ELECTION chaos in November. Retrieved February 16, 2021, from <https://www.nytimes.com/2020/09/03/technology/facebook-election-chaos-november.html>

Jigsaw. (n.d.). Retrieved February 16, 2021, from <https://jigsaw.google.com>

Kertysova, K. (2018). Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered. *Security and Human Rights*, 29(1-4), 55–81. <https://doi.org/10.1163/18750230-02901005>

Kleinman, Z. (2018, March 09). Fake news 'travels faster', study finds. Retrieved February 16, 2021, from <https://www.bbc.com/news/technology-43344256>

Lazer, B. (2018). The science of fake news. *Science (American Association for the Advancement of Science)*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

Lazer, D., Baum, M., Benkler, Y., Berinsky, A., Greenhill, K., Menczer, F., ... Zittrain, J. (2018, March 09). The science of fake news. Retrieved February 16, 2021, from <https://science.sciencemag.org/content/359/6380/1094>



Let's flatten the infodemic curve. (n.d.). Retrieved February 16, 2021, from <https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve>

Lies, propaganda and fake news: A challenge for our age. (n.d.). Retrieved February 16, 2021, from <https://www.bbc.com/future/article/20170301-lies-propaganda-and-fake-news-a-grand-challenge-of-our-age>

Macheso, E., & Kingori, D. (2021, January 14). #Defyhatenow_South sudan. Retrieved February 16, 2021, from https://defyhatenow.org/defyhatenow_southsudan

Managing the COVID-19 INFODEMIC: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. (2020, September 23). Retrieved February 16, 2021, from <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>

Martinez, M. (2018, November 12). Burned to death because of a rumour on WhatsApp. Retrieved February 16, 2021, from <https://www.bbc.com/news/world-latin-america-46145986>

McNaughton, J. (2016, November 20). A moment of truth for Mark Zuckerberg. Retrieved February 16, 2021, from <https://www.theguardian.com/commentisfree/2016/nov/20/fake-news-facebook-mark-zuckerberg-donald-trump>

Medical activities suspended after Ebola treatment CENTRE attack: MSF. (2019, February 26). Retrieved February 16, 2021, from <https://www.msf.org/medical-activities-suspended-after-ebola-treatment-centre-attack>

Michael, M. (2020, June 09). Yemen's rebels crack down as COVID-19 and rumors spread. Retrieved February 16, 2021, from <https://apnews.com/article/677a1fc12d864cd37eea57e5f71614a2>

MIT Sloan Office of Media Relations. (2019, January 28). MIT Sloan study finds crowdsourcing an effective tool to fight spread of fake news. Retrieved February 16, 2021, from <https://mitsloan.mit.edu/press/mit-sloan-study-finds-crowdsourcing-effective-tool-to-fight-spread-fake-news>

Mondal, M., Silva, L. A., & Benevenuto, F. (2017). A measurement study of hate speech in social media. Proceedings of the 28th ACM Conference on Hypertext and Social Media. doi:10.1145/3078714.3078723

Mondal, S. (2018). Characterizing usage of explicit hate expressions in social media. The New Review of Hypermedia and Multimedia, 24(2), 110–130. <https://doi.org/10.1080/13614568.2018.1489001>

Montalto Monella, L. (2020, March 03). Orban uses coronavirus as excuse to suspend asylum rights in Hungary. Retrieved February 16, 2021, from <https://www.euronews.com/2020/03/03/orban-uses-coronavirus-as-excuse-to-suspend-asylum-rights-in-hungary>

Mozilla foundation - Internet Health Report. (2019). Retrieved February 16, 2021, from <https://foundation.mozilla.org/en/insights/internet-health-report>

Myanmar: UN fact-finding Mission releases its full account of massive violations by military in Rakhine, Kachin and Shan States. (2018, September 18). Retrieved March 01, 2021, from <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=23575&LangID=E>



NATO. (2018). Brussels summit Declaration issued by the heads of state and government participating in the meeting of the North Atlantic Council in Brussels, 11-12 July 2018. Retrieved February 16, 2021, from https://www.nato.int/cps/en/natohq/official_texts_156624.htm

NATO. (n.d.). London declaration. Retrieved February 16, 2021, from https://www.nato.int/cps/en/natohq/official_texts_171584.htm

Nduka Orjinmo & Ajoke Ulohotse. (2020, April 15). Lagos unrest: The mystery of Nigeria's Fake Gangster attacks. Retrieved February 16, 2021, from <https://www.bbc.com/news/world-africa-52285271>

Nemr, C., Gangware, W., & Park Advisors. (2019). Weapons of Mass distraction. Retrieved February 16, 2021, from <https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distracton-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf>

Network Contagion Research Institute. (2021, January 13). Retrieved February 16, 2021, from https://networkcontagion.us/?fbclid=IwAR0ymxsbl_V4F_8KTISe5UhS2uHbglisa7toaS3fAna8KPNpfbqZGA3OSOI

New national security unit set up to tackle fake news in the UK. (2018, January 23). Retrieved February 16, 2021, from <https://www.theguardian.com/politics/2018/jan/23/new-national-security-unit-will-tackle-spread-of-fake-news-in-uk>

News literacy project. (2021, February 12). Retrieved February 16, 2021, from <https://newsliit.org>

Nigerian police say "fake news" on Facebook is killing people. (n.d.). Retrieved February 16, 2021, from https://www.bbc.co.uk/news/resources/idx-sh/nigeria_fake_news

Russia's support for Trump's election is no longer disputable. (2018, December 18). Retrieved February 16, 2021, from https://www.washingtonpost.com/opinions/russias-support-for-trumps-election-is-no-longer-disputable/2018/12/17/d194870e-023a-11e9-9122-82e98f91ee6f_story.html

O'Connor, G., & Schneider, A. (2017, April 03). How Russian twitter Bots pumped out fake news during the 2016 election. NPR. Retrieved March 22, 2021, from <https://www.npr.org/sections/alltechconsidered/2017/04/03/522503844/how-russian-twitter-bots-pumped-out-fake-news-during-the-2016-election>

Ortutay, B. (2020, November 06). Facebook bans big 'stop The steal' group for SOWING violence. Retrieved February 16, 2021, from <https://apnews.com/article/election-2020-donald-trump-misinformation-violence-elections-d5c9bd5fe6a799fd627c50521b6cbb36>

Park Hong, C. (2020, December 20). How fake news is hatching in immigrant communities. Retrieved February 16, 2021, from <https://www.nytimes.com/2020/12/20/opinion/fake-news-disinformation-immigrants.html>

Parks, M., & Dougliis, S. (2019, October 31). Fake news: How to spot misinformation. Retrieved February 16, 2021, from <https://www.npr.org/2019/10/29/774541010/fake-news-is-scary-heres-how-to-spot-misinformation>

Pennycook, G., Bear, A., Collins, E., & Rand, D. (2017, September 14). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. Retrieved February 16, 2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035384

Pennycook, R. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences - PNAS, 116(7), 2521–2526. <https://doi.org/10.1073/pnas.1806781116>



Permanent suspension of @realdonaldtrump. (n.d.). Retrieved February 16, 2021, from https://blog.twitter.com/en_us/topics/company/2020/suspension.html

PolitiFact - the principles of The TRUTH-O-METER: PolitiFact's methodology for independent fact-checking. (n.d.). Retrieved February 16, 2021, from <https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/#Truth-O-Meter%20ratings>

Polyakova, A., & Meserole, C. (2019, November 25). Exporting digital authoritarianism. Retrieved February 16, 2021, from <https://www.brookings.edu/research/exporting-digital-authoritarianism>

Ponniah, K. (2019, April 05). WhatsApp: The 'black hole' of fake news in India's election. BBC News. Retrieved February 16, 2021, from <https://www.bbc.com/news/world-asia-india-47797151>

Potkin, F., & McPherson, P. (2020, November 06). 'Spreading like wildfire': Facebook fights hate speech, misinformation before Myanmar Poll. Reuters. Retrieved February 16, 2021, from <https://uk.reuters.com/article/us-myanmar-election-facebook-idUKKBN27L300>

Project origin. (n.d.). Retrieved February 16, 2021, from <https://www.originproject.info/about>

Purnell, N. (2019, April 13). WhatsApp Users Spread Antivaccine Rumors in India. Retrieved February 16, 2021, from <https://www.wsj.com/articles/whatsapp-users-spread-antivaccine-rumors-in-india-11555153203>

Reed, J. (2018, February 22). Hate speech, atrocities and fake news: The crisis in Myanmar. Retrieved February 16, 2021, from <https://www.ft.com/content/2003d54e-169a-11e8-9376-4a6390addb44>

Resist counter disinformation toolkit. (2020, November 10). Retrieved February 16, 2021, from <https://gcs.civilservice.gov.uk/publications/resist-counter-disinformation-toolkit>

Roose, K. (2017, October 29). Forget Washington. Facebook's Problems abroad are far more disturbing. Retrieved February 16, 2021, from <https://www.nytimes.com/2017/10/29/business/facebook-misinformation-abroad.html>

Security and Intelligence Threats to Elections (SITE) Task Force. (2019, February 07). Retrieved February 16, 2021, from <https://www.canada.ca/en/democratic-institutions/services/protecting-democracy/security-task-force.html>

Send in the trolls: CANADA braces for an online disinformation assault On COVID-19 | CBC News. (2020, March 14). Retrieved February 16, 2021, from <https://www.cbc.ca/news/politics/covid-19-coronavirus-disinformation-trolls-1.5497805>

Services: Fact CHECKING, Media intelligence & ANTI-TERRORISM Tools. (n.d.). Retrieved February 16, 2021, from <https://www.logically.ai/services>

Share verified. (n.d.). Retrieved February 16, 2021, from <https://shareverified.com/en>

Shu, B. (2020). Combating disinformation in a social media age. Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 10(6). <https://doi.org/10.1002/widm.1385>

Slick. (n.d.). Commit to transparency - sign up for the International Fact-checking network's code of principles. Retrieved February 16, 2021, from <https://ifcncodeofprinciples.poynter.org>



So, you think you've spotted some 'fake news' - Now What? | CBC News. (2019, July 05). Retrieved February 16, 2021, from <https://www.cbc.ca/news/technology/fake-news-disinformation-propaganda-internet-1.5196964>

South Sudan continues to Face PERSISTENT challenges to peace and stability, Security council told | | UN NEWS. (n.d.). Retrieved February 16, 2021, from <https://news.un.org/en/story/2016/11/545682-south-sudan-continues-face-persistent-challenges-peace-and-stability-security#.WNIZWhD6KCY>

Global News. (2020, December 04). CSIS accuses Russia, China and Iran of spreading COVID-19 disinformation. Retrieved February 16, 2021, from <https://globalnews.ca/news/7494689/csis-accuses-russia-china-iran-coronavirus-covid-19-disinformation>

Subramanian, S. (2020, October 14). The Macedonian teens who MASTERED fake news. Retrieved February 16, 2021, from <https://www.wired.com/2017/02/veles-macedonia-fake-news>

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769. doi:10.1111/j.1540-5907.2006.00214.x

Taub, A., & Fisher, M. (2018, April 21). Where countries Are TINDERBOXES and Facebook is a match. *The New York Times*. Retrieved February 16, 2021, from <https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html>

Tavernise, S. (2016, December 07). As fake news spreads lies, more readers shrug at the truth. Retrieved February 16, 2021, from <https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html>

The Accountability Lab - About us. (2020, March 12). Retrieved February 16, 2021, from <https://accountabilitylab.org/about-us>

The godfather of fake news. (n.d.). Retrieved February 16, 2021, from https://www.bbc.co.uk/news/resources/idt-sh/the-godfather_of_fake_news

The International Rescue Committee warns against worrying rise of misinformation in crisis-affected States, IMPERILING fight against COVID-19. (2020, October 28). Retrieved February 16, 2021, from <https://www.rescue.org/press-release/international-rescue-committee-warns-against-worrying-rise-misinformation-crisis>

The rise of digital authoritarianism. (2018). Retrieved February 16, 2021, from <https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism>

Together we Are Stronger: Social media companies, civil society, and the fight against disinformation. (n.d.). Retrieved February 16, 2021, from <https://freedomhouse.org/article/together-we-are-stronger-social-media-companies-civil-society-and-fight-against>

Tools that fight disinformation online. (n.d.). Retrieved February 16, 2021, from <https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>

Twitter bans David Icke over Covid misinformation. (2020, November 04). Retrieved February 16, 2021, from <https://www.bbc.com/news/technology-54804240>

UN tackles 'infodemic' of misinformation and cybercrime IN COVID-19 Crisis. (n.d.). Retrieved February 16, 2021, from <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19>



United Nations official document. (2016). Retrieved February 16, 2021, from https://www.un.org/ga/search/view_doc.asp?symbol=S%2F2016%2F963

Updating our rules against hateful conduct. (2019). Retrieved February 16, 2021, from https://blog.twitter.com/en_us/topics/company/2019/hatefulconductupdate.html

Verma R. (2020). China's diplomacy and changing the COVID-19 narrative. *International Journal*, 75(2):248-258. doi:10.1177/0020702020930054

Vicario, Q. (2019). Polarization and Fake News: Early Warning of Potential Misinformation Targets. *ACM Transactions on the Web*, 13(2), 1–22. <https://doi.org/10.1145/3316809>

Vosoughi, A., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science (American Association for the Advancement of Science)*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

Wanjala, E. (2017). Red cross alarmed by prank calls on election violence. Retrieved February 16, 2021, from <https://www.the-star.co.ke/news/2017-08-14-red-cross-alarmed-by-prank-calls-on-election-violence>

Wardle, C., & Derakhshan, H. (n.d.). Information disorder: Toward an interdisciplinary framework for research. Retrieved February 16, 2021, from <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>

What is an algorithm? (2020, December 09). Retrieved February 16, 2021, from <https://www.bbc.co.uk/bitesize/topics/z3tbwmn/articles/z3whpv4>

WhatsApp to impose new limit on forwarding to fight fake news. (2020, April 07). Retrieved February 16, 2021, from <https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news>

Wheeler, T. (2017, November 14). Did technology kill the truth? Retrieved February 16, 2021, from <https://www.brookings.edu/research/did-technology-kill-the-truth>

Woolley, S. C., & Howard, P. N. (2017). Computational propaganda worldwide: Executive summary. Retrieved March 22, 2021, from <https://demtech.oii.ox.ac.uk/research/posts/computational-propaganda-worldwide-executive-summary>

WHO health Alert brings COVID-19 facts to billions via WhatsApp. (2020). Retrieved February 16, 2021, from <https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp>

Who said what? The Security Challenges of Modern Disinformation. (2018). Retrieved February 16, 2021, from https://www.canada.ca/content/dam/csis-scrs/documents/publications/disinformation_post-report_eng.pdf

Why Twitter is the best social media platform for disinformation. (n.d.). Retrieved February 16, 2021, from <https://www.vice.com/en/article/bj7vam/why-twitter-is-the-best-social-media-platform-for-disinformation>

Woolley, S. (2020, April 02). We're fighting fake news AI bots by using more AI. That's a mistake. Retrieved February 16, 2021, from <https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake>

Yayboke, E., & Brannen, S. (2021, February 08). Promote and build: A strategic approach to digital authoritarianism. Retrieved February 16, 2021, from <https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism>