Understanding and Countering Gendered Disinformation

A framework for resilience and action

May 2025

CSKA Community Safety Knowledge Alliance

SAPPER LABS

This page is intentionally left blank

This work was prepared by:

www.cskacanada.ca www.sapperlabs.com

With funding from:

Canada

Authors of report

Janos Botschner, PhD
Giovanna Cioffi, CD, PhD
Dave McMahon, MSM, BEng
Julie Ollinger, PhD
Bradley Sylvestre, MA
Ritesh Kotak, JD
Cal Corley, MBA

Additional contributors

Additional support to this project was provided by Actua (www.actua.ca). The following individuals co-authored knowledge products for parents, youth and educators in collaboration with the project team. These resources were produced by Actua in partnership with CSKA:

Janelle Fournier, PhD (ABD)
Mikayla Ellis, BA
Abbey Ramdeo, MT

Suggested report citation:

Botschner, J., Cioffi, G., McMahon, D., Ollinger, J., Sylvestre, B., Kotak, R. & Corley, C. (2025). Understanding and countering gendered disinformation: A framework for resilience and action. Ottawa ON: Community Safety Knowledge Alliance.

Correspondence:

Jbotschner[at]cskacanada.ca

About the Community Safety Knowledge Alliance

The Community Safety Knowledge Alliance is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes.

Over the past decade, CSKA has conducted interdisciplinary research on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

About Sapper Labs Group

Sapper Labs Group conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network.

The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

ACKNOWLEDGEMENTS

The authors would like to express sincere appreciation to the following individuals/groups for the guidance they provided to this work. The contents, conclusions and recommendations are those of the authors, alone.

Project Advisory Committee

Michael Doucet, former Executive Director, National Security Intelligence Review Agency
Jennifer Flanagan, Chief Executive Officer, Actua
Carmen Gill, Professor, Department of Sociology, University of New Brunswick
Jennifer Irish, Director, Information Integrity Lab, University of Ottawa
Alan Jones, Executive Advisor, Professional Development Institute, University of Ottawa; former Assistant Director, Canadian Security Intelligence Service
Marcus Kolga, Founder and Director, DisinfoWatch; Fellow, MacDonald-Laurier and Conference of Defence Associations Institutes

Development of Knowledge Products

For parents and educators

Actual National STEM Educator Community of Practice
Actua staff and Actua Network members

For youth

Actua National Black Youth in STEM Program Youth Delegation
Actua Indigenous Youth in STEM Program Youth Delegation
Actua staff and Actua Network members

For police and community groups

Delta Police Department

André Cruz
Communications Assoc.

Cst. Derek Defrane
Domestic Violence Unit

Kim Gramlich
Mgr., Victim Services

Sgt. Alex Quezada
Vulnerable Sector Unit

Greater Sudbury Police Service

Det. Sgt. Adam Demers
Criminal Investigations/ Intimate Partner Violence

Dan Gelinas
Community Mobilization Liaison

Det. Sgt. Lee Rinaldi
Major Sex Crimes

Sudbury YWCA

Marlene Gorman
Executive Director

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Today's interconnected world provides a wealth of opportunities for those wishing to harm women, girls and gender diverse persons individually and at scale. This report describes the mechanisms, impacts, and actors behind technology-enabled gendered disinformation. This is not just a gender issue – it is also a socioeconomic and public safety issue which, in some cases, may also become a national security concern. We illustrate why action is needed now and chart a theory- and evidence-informed path forward.

Technology-enabled gender-based violence – including disinformation – draws from a powerful arsenal of tools. It can be used for illicit surveillance (such as monitoring movement and communications) and to manipulate aspects of the built environment (such as features of "smart" homes and vehicles). It can also be used to pollute the information space with deceptive narratives. Gendered disinformation poses a dual threat: it endangers individuals, especially women and gender-diverse people who are often its direct targets; and it undermines society by eroding trust and cohesion, silencing voices, and weakening democratic norms and processes.

Members of certain populations – notably, marginalized and racialized women, girls and gender diverse persons – may disproportionately encounter greater levels of gendered disinformation. Indigenous women and girls in Canada often face gendered disinformation and related violence due to historical biases, colonial legacies and contemporary social media narratives that can often perpetuate harm.

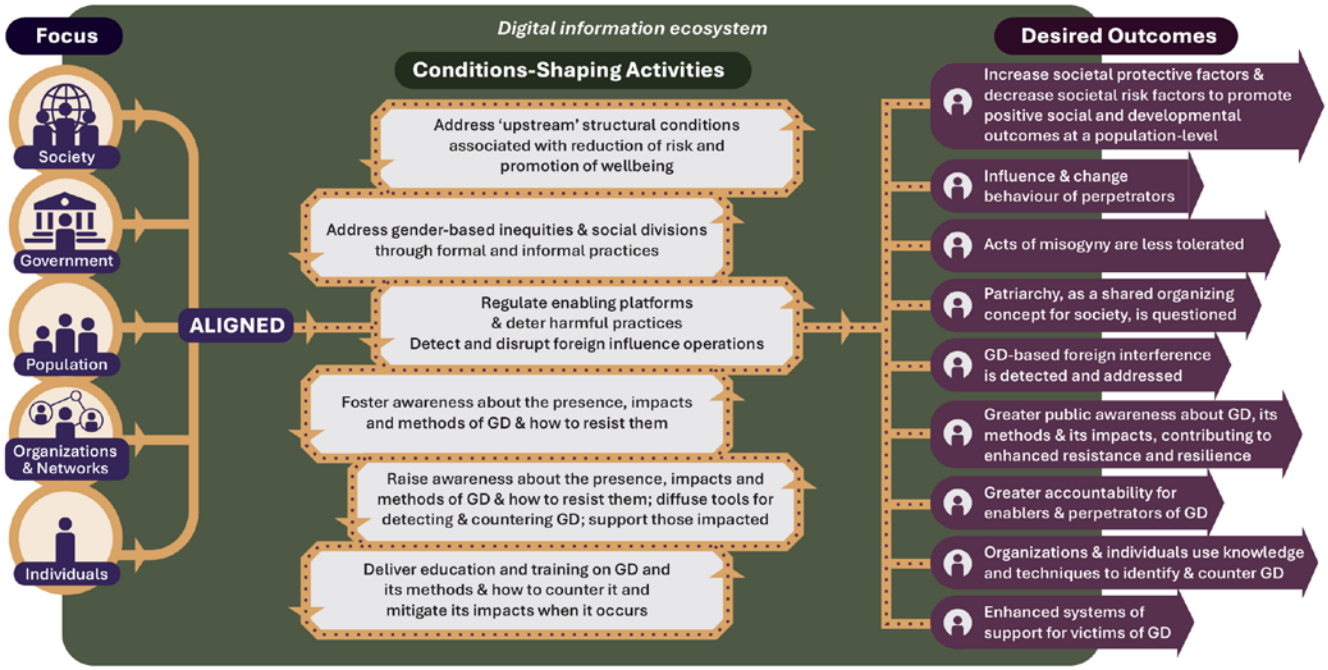
Gendered disinformation is not spread by chance. It can be driven by individual actors, aligned domestic and transnational ideological groups, and even nation states that seek to destabilize democratic societies. When foreign governments are involved, GD becomes an instrument used to sow division, fear, and mistrust across borders – sometimes as a component of broader influence or cyber operations. At a time when online spaces too often amplify misogynist voices and targeted abuse, understanding and countering gendered disinformation has never been more urgent. It is a shared threat across society. Consequently, the resolve and the ability to address gendered disinformation must be a matter of shared responsibility.

The widespread occurrence of gendered disinformation around the world, often leading to violence, underscores the need for international cooperation. This is essential to address the complex, cross-border nature of the issue effectively. By sharing best practices, resources, and intelligence, countries can develop unified strategies to combat disinformation. Domestically, integrating these global insights into national policies and practices will enhance local efforts, ensuring that responses are comprehensive and culturally relevant. Joint initiatives can also strengthen diplomatic relations, promote gender equality, and uphold human rights on a broader scale.

This report provides a novel perspective on gendered disinformation, including a framework for action with a corresponding system of people, processes and technology. Furthermore, it provides a short set of pragmatic recommendations that will have significant impact on combatting gendered disinformation, enhancing human rights protection, and promoting gender equality. These elements are accompanied by a set of information resources for key stakeholder groups seeking to raise awareness and to counter this complex, multi-layered problem.

Building the capacity to counter gendered disinformation will require collaboration. As we navigate geo-political and domestic tensions that threaten the cohesion, unity and sovereignty of Canadian society, our willingness to confront and respond to gendered disinformation will shape the resilience and inclusivity of our digital, democratic and social spaces for years to come.

A preliminary theory of change for addressing gendered disinformation is depicted below. This is discussed in further detail at Figure 9 in this report.



This theory involves multi-level efforts designed to align and create mutually reinforcing conditions, significantly enhancing the likelihood of achieving a range of desired outcomes.

Conclusions

Addressing gendered disinformation is crucial for safeguarding human rights, promoting gender equality, and upholding democratic values. This issue, intertwined with polarization, patriarchy, and misogyny, targets women, girls, and gender-nonconforming individuals, causing harm. A strategic, multi-layered approach is necessary to combat this, focusing on awareness and a coordinated

response. Strengthening resistance to such disinformation requires collaborative efforts to prevent risks, enhance resilience, and align solutions with democratic principles.

Gendered disinformation about Indigenous women and girls in Canada is exacerbated by a colonial history that persists today, reinforcing harmful stereotypes and ignoring ongoing violence. Multi-faceted efforts must be undertaken to break this cycle by challenging false narratives, reforming media practices, and prioritizing Indigenous voices in storytelling. Such measures are vital for transforming the information landscape and supporting reconciliation.

The path forward emphasizes multi-sector collaboration and building broad-based networked capacity to counter gendered disinformation. Increasing awareness and developing new knowledge will be central to this effort. This approach should foster mutual benefits and support collective learning, planning, implementation and further research.

We propose a comprehensive theory of change involving strategically aligned, society-wide interventions grounded in the leading research. This approach includes providing a robust set of knowledge resources and technology examples beneficial to professionals in human services, policy-making, and national security. Furthermore, we recommend creating a cross-sectoral network dedicated to knowledge development and mobilization. This network will support evidence-based, collaborative efforts, ensuring that interventions are informed by the best available evidence and practices. By fostering cooperation across multiple sectors, this critical issue can be tackled effectively and holistically.

Recommendations:

Policy, Legislation and Enforcement

1. That the federal government:

    a. Implement policy and legislative measures to counter gendered disinformation, recognizing that it is a threat that spans community safety and wellbeing, and national security.
        - *The corresponding regulatory framework should ensure platform accountability, transparency, and meaningful financial penalties for non-compliance.*

    b. With targeted investment, initiate cross-departmental, industry, academic and private sector operational coordination and program collaboration to address gendered disinformation within public safety, public health, digital regulation, defence and national security frameworks.

c. Develop a national strategy on gendered disinformation in close partnership with the private sector, research and civil society, integrating public safety, digital governance, and foreign policy approaches.

d. Convene and engage women's advocacy organizations, racial justice groups, security and intelligence professionals, academic researchers, cyber-security experts and relevant community and private sector entities in dialogue on such matters as how to optimize the balance of protection and enforcement with freedom of expression online.

e. Increase data collection and monitoring of gendered disinformation trends and actionable current intelligence.

f. Conduct periodic cross-sector consultations with experts in gender-based violence, cybersecurity, open source intelligence, national security, and digital regulation to understand the evolving landscape of gendered disinformation.

g. Establish gender-responsive online safety laws that hold technology platforms accountable. Options include the re-introduction of Bill C-36 and the applications of relevant elements of a Clean Pipes Strategy.

h. Enhance training for security, intelligence, diplomatic, defence, law-enforcement and policymakers on technology-enabled GD.

i. Invest in digital literacy, research, open source intelligence and enforcement mechanisms to strengthen Canada's resilience against gendered disinformation.

Research and Knowledge Mobilization

2. That the Government of Canada support the creation of a cross-sectoral knowledge mobilization network on gendered disinformation – the Gendered Disinformation Knowledge Network (GenD-Net).

   Such a network would serve as a hub for leadership, information sharing, education and training, research, and policy coordination, program planning, operational coordination and de-confliction ensuring that responses to gendered disinformation are evidence-based, and aligned across sectors.

   The objectives of the network will be to:

   - *Enhance knowledge mobilization and public awareness of gendered disinformation.*
   - *Support curriculum development, stimulate and contribute to education and training.*

- *Strengthen community and cross-sectoral dialogue and collaboration on policy development.*
- *Support defence, intelligence, police and public safety agencies.*
- *Advance research and innovation, including evaluation capacity building.*
- *Bridge gaps in service provision for affected communities.*

Gendered Disinformation as a National Security Issue

3. That the Government of Canada refine and implement options for countering gendered disinformation as a national security issue, including its use as an element of foreign interference. Enhance the capabilities of defensive cyber operations in relation to this threat. More particularly:
    a. Establish a dedicated government funding stream for research and innovation on gendered disinformation that is open to Canadian industry, academia and not-for profit organizations.
    b. Incentivize Canadian industry participation and innovation through public-private partnerships and direct investment.
    c. Develop a national strategy on gendered disinformation as a foreign interference threat, and ensure integration with national defence policy, cyber security and national security strategies.
    d. Fund the creation of a cross-sectoral intelligence-sharing network to combat gendered disinformation, including the creation and maintenance of a national gendered disinformation threat landscape reporting capacity; this would, in-turn, feed into an intelligence "dashboard" (Figure 11) which could be made publicly available as part of building overall awareness an public will to confront this problem (See Annex E4, Attachment B).
    e. Establish legal and policy frameworks to protect women in public life from both foreign and domestic online harm.
    f. Develop a rapid response mechanism to protect individuals facing high-risk disinformation attacks (see Annex E4, Briefing Resources 1 and 4).

Impact of Recommendations

Implementing these recommendations will have significant impacts on combatting gendered disinformation, enhancing human rights protection, and promoting gender equality. By addressing this issue, intertwined with polarization and misogyny, we can safeguard women, girls, and gender-nonconforming individuals from targeted harm. More specific areas impacted are as follows:

Policy and Legislation

By implementing comprehensive policies and legislation, the federal government will strengthen community safety and national security. Establishing regulatory frameworks with platform accountability and penalties for non-compliance will ensure that digital spaces are safer and more transparent. Cross-departmental coordination will enhance efforts to address gendered disinformation within public safety and national security frameworks.

Multi-Sector Collaboration

Creating a national strategy in partnership with the private sector, research institutions and civil society will integrate approaches to enhancing both public safety and social media governance. Engaging diverse organizations in dialogue will balance safety and security with freedom of expression. Furthermore, this approach will help build resilience against gendered disinformation through enhanced data collection, training, and digital literacy investments.

Research and Knowledge Mobilization

A dedicated funding stream for research and innovation, alongside public-private partnerships, will drive industry participation and technological advancements.

Establishing the Gendered Disinformation Knowledge Network (GenD-Net) will enhance public awareness, support curriculum development, and foster cross-sectoral collaboration. By bridging gaps in service provision, it will ensure evidence-based responses aligned across sectors.

National Security

Recognizing gendered disinformation as a national security issue will help refine strategies to counter foreign interference. Developing a rapid response mechanism and legal frameworks will protect individuals from high-risk disinformation attacks.

Overall, when implemented, these measures will help to transform the online information landscape, support reconciliation, and uphold Canadian liberal democratic values by fostering a coordinated, strategic response to gendered disinformation.

COMMON TERMS AND TECHNIQUES

Misinformation is untrue content that is spread by people who believe that it is true – ***untrue information, good or neutral intent***. **Disinformation** is untrue content that is spread by people who know that it is untrue. Misinformation could be spread innocently, or to cause harm. Disinformation is always spread knowingly and deliberately to cause harm – ***untrue information, bad intent***. **Malinformation** is information that is true, but it's shared in a way that's meant to cause harm – ***true information, bad intent***.

Fake stories – Fake news articles or social media posts that attack individuals, such as former partners/spouses, or those in public or leadership roles.

Non-consensual image sharing – Can include posting or re-posting intimate images that were meant to be private or exclusive to a partner. It can also involve uploading sexual photos or videos of an ex-partner to social media or pornographic websites without their consent.

Manipulated images & videos – Edited pictures or **"deepfake"** videos that make it look like someone said or did something they never did. Commonly encountered situations include non-consensual, out-of-context, sharing of manipulated or real photos/fake explicit content.

Misinformation about gender roles – Posts or comments claiming that women are naturally bad at leadership, science, or sports.

Harassment & cyberbullying – Online attacks that try to intimidate, humiliate, or silence.

Fake accounts & impersonation – Creating fake online profiles to spread lies, harass someone, or damage their reputation. When this involves creating one or many fake accounts, or taking over existing accounts to make it look like people agree with a fake story, it is called **astroturfing**.

Memes & satire – These are jokes or cartoons that disguise harmful messages about their targets as "just humour."

Doxxing – Broadly sharing private information (like a home address or phone number) online to intimidate or harm a person. Sometimes, this can lead to offline intimidation or violence.

Surveillance and manipulation of "smart" technology – For example, using commercially available tracking devices, to monitor someone's movement; or manipulating home or vehicle systems to intimidate someone.

Cyberflashing is the act of sending someone unsolicited sexual images through digital means, often via text message, social media, dating apps, or file-sharing features like AirDrop or Bluetooth.

Catfishing – a practice in which individuals create fake online identities to deceive others, often for abusive or exploitative purposes.

INTRODUCTION

Gender equality and the safety of women, girls and gender non-conforming persons are under threat. Ideological movements, political forces, and global tensions rooted in patriarchy and misogyny are deepening social and political divides online. In some cases, these divisions are intentionally exploited to harm individuals and destabilize Canadian society.

The aim of this research and development project was to create a framework and a set of corresponding practices to understand and counter online gendered disinformation. This issue occurs against a more general backdrop of: technology-enabled gender-based violence and repression[1], which is a global problem; and foreign interference, which has been flagged as a significant and ongoing threat to Canada and to Canadians (Public Inquiry into Foreign Interference in Federal Electoral Processes and Democratic Institutions, 2025).

Generative AI is driving a surge in disinformation. One specific category – gendered disinformation (GD) – targets women, girls and gender-diverse persons.[2] – through misogynistic harassment and intimate partner violence. Furthermore, it can collectively deny them a voice and undermine their role in society.

Gendered disinformation makes use of distinctions and divisions centred on traditional notions of gender. It targets those who do not conform, such as working women, transgender individuals, or those with non-traditional gender expressions. It targets specific groups to undermine social cohesion and public trust (e.g., by questioning the competence of female leaders or the appropriateness of certain books available in school libraries). GD leverages digital technologies, especially social media, to amplify impact. Finally, GD not only harms individuals, but it

The use of tropes or memes to promote gendered disinformation.

A **trope** is a commonly used theme, idea, or storytelling device that helps people quickly understand a situation or character. They can be visual, verbal, or conceptual, and they often rely on familiar patterns. When used for disinformation, they may reinforce stereotypes.

Example: *"Women are bad drivers."*

A **meme** is a piece of content – often an image, video, or phrase – that spreads quickly online and is shared, adapted, and remixed by different people. They can be humorous, political, or cultural, and they often carry deeper meaning in a short, relatable format. In online disinformation, memes are used to spread false or harmful messages in a way that feels casual and shareable, making them powerful tools for manipulation

Example: *"Real women ['trad wives'] don't chase careers—they support their husbands and raise children the right way."*

[1] Also known as online violence against women.

[2] For consistency, we use the term, "women and gender diverse", to refer to individuals – such as women, gender non-conforming persons, and those with various sexual identities –who are targeted by gendered disinformation or used as instruments for malicious purposes.

deepens existing social divisions, making it more difficult to achieve equality and mutual understanding.

The implications of gendered disinformation are both deeply personal and profoundly political. Targeted individuals may experience psychological distress and reputational damage. These harms can deter affected individuals from participating in politics, activism, or journalism, ultimately silencing important voices in public discourse. At a broader level, the spread of online gendered disinformation erodes public trust in media and institutions – particularly when false content is mistaken for real, or genuine content is dismissed as fabricated. Moreover, disinformation campaigns frequently reinforce harmful narratives, perpetuating stereotypes that disproportionately impact women, girls and LGBTQIA+ communities (e.g., Richardson-Self, 2021; Sobieraj, 2020).

Contested truth and false or misleading content (e.g., text, audio, video and images) scaled through social media is one of the most serious threats to democratic values, civic participation, domestic tranquility, and the ability of people and nations to collaborate in addressing the complex challenges of this century.

Recent examples identified and discussed in print and media publications include: videos of anti-woman influencers denouncing female empowerment, including one posted by "manosphere" influencer Andrew Tate opining that women should not be allowed to drive (Donegan, 2025); the role of video gaming platforms in spreading misogynistic tropes and memes (Stuart, 2025); accounts of how disinformation campaigns can be used to destroy the reputations of political opponents (Ressa, 2022); and the weaponization of the term "woke" to attack various initiatives focused on inclusion (Off, 2024).

In the United States, the National Democratic Institute (2022) described GD as a critical issue for democracies because of its impacts on the participation of women in online political activity, and in recognition of its use by authoritarian and illiberal actors as a tactic of online violence aimed at silencing and undermining the political agency of women and girls. Thus, an increase in GD benefits individual perpetrators of information-based violence and groups seeking to disrupt social harmony and undermine the value of inclusion.

> *A healthy democracy is characterized by a vibrant and diverse range of voices and groups, engaged in a constant process of deliberation, discussion, negotiation and compromise. Because of this characteristic, democracy requires a civic setting in which people can freely express their ideas. To create such a setting, democracy relies on values and principles such as the equality of individuals and respect for others, as well as consideration for the diversity of opinions and beliefs. It requires social and political institutions that encourage the participation of all. By fostering distrust, creating division and preventing compromise, disinformation threatens this fundamental feature of democracy.*
>
> (Public Inquiry Into Foreign Interference in Federal Electoral Processes and Democratic Institutions, 2025).

In recent testimony before the House of Commons Standing Committee on Public Safety and National Security, Marcus Kolga, of DisInfoWatch, argued that "[s]afeguarding Canada's cognitive sovereignty and the integrity of our information environment is essential to defending our democracy and maintaining social cohesion" (Parliament of Canada, 2024).

In the final report of the Public Inquiry Into Foreign Interference in Federal Electoral Processes and Democratic Institutions, Commissioner Hogue opined that disinformation is a pronounced threat to Canadian society.

Technology-Enabled Violence Against Women is a Widespread Problem That Violates Human Rights

The September 2024 SDG Gender Index, published by Equal Measures 2030 (a global coalition of NGOs that use data and evidence to address gender equality), determined that none of the 139 countries assessed has achieved the UN's 2030 SDG benchmarks for gender equality. While Canada is ranked 18th and falls in the 'good' category[3], its progress has stalled over the recent measurement periods (2015-2019, 2019-2022) and is projected to remain unchanged from 2022 to 2030.

Moreover, the report stated that gender-based violence against Indigenous women and girls is a particularly serious issue in Canada. These individuals face greater levels of both intimate and non-intimate partner violence compared to non-Indigenous females. A recent study by the Canadian Women's Foundation found that thirty percent of Indigenous women encounter unwelcome behaviour, including online[4]. The study also found that found one in five Canadian women experiences some form of online harassment.

The increased risk of online abuse by Indigenous women reflects long-standing colonial practices of objectification, sexualization and misrepresentation (Corbett, 2019). These assumptions and biases provide a foundation for contemporary narratives in media and popular culture that perpetuate harm. For example, stereotypes and disinformation about missing and murdered Indigenous women and girls distract from many of the deeper causes of this violence (Corbett, 2019). This can lead to misunderstanding and apathy, undermining public commitment for meaningful change.

[3] Following Belgium, ahead of Spain and the United Kingdom and the United States.
[4] Canadian Women's Foundation (n.d.)

A recent report by the Institute of Global Politics and the Vital Voices Global Partnership (Jankowicz, et al., 2024) on technology-enabled gender-based violence[5] confirms that online abuse of women is a widespread problem across all continents. The associated global statistics[6] are stark: between 2019 and 2020, 85 percent of women had witnessed or experienced online gender-based violence and 38 percent had been personally impacted by it. The Economist Intelligence Unit (EIU, 2020) assessed that these figures likely underestimate the actual prevalence of the issue.

> *The repetitive representation of Indigenous women engaging in "high-risk" lifestyles normalizes the violence against them....*
>
> *The silencing of violence against Indigenous women and girls is made worse in comparison to the media's compassionate framing of white women.*
>
> (Corbett, 2019)

The Economist Intelligence Unit reported a North American prevalence of online gender-based violence of 76 percent[7]. In a University of Maryland study reported by Hess (2014), created a set of fake online accounts with feminine and masculine usernames. They then distributed them across various online chat rooms. Accounts with feminine usernames received an average of 100 sexually explicit or threatening messages per day, whereas those with masculine usernames received only 3.7 such messages.

Sobieraj (2020) suggests that the uneven distribution of identity-based abuse among women is linked to power and inequality. She observes that online attacks are most severe for three groups: women with multiple marginalized identities; those who publicly critique male-dominated spaces; and those perceived as feminist or non-conforming with traditional gender norms. Women at the intersection of all three groups, such as BIPOC[8] feminist members of the LGBTQIA+ community, may be particularly targeted. Researchers studying far-right extremist movements have observed that "persistent anxiety about masculinity" is a core feature of these ideologies (Kesevan, 2024).

According to the EIU (2020) study, nine threat tactics predominated across respondents to its online survey:

- Misinformation and defamation (67 percent);
- Hate speech (65 percent) and violent threats (52 percent);
- Cyber harassment (66 percent), hacking and stalking (63 percent);
- Doxing[9] (55 percent);
- Astroturfing[10] (58) percent;

[5] Known as technology-facilitated violence against women (TF-VAW) in the research literature
[6] Economist Intelligence Unit (2020) data from 2019-2020, reported by Jankowicz, et al. (2024)
[7] While this figure is high, it is the second-lowest across continents, with Europe being the lowest at 74 percent.
[8] i.e., Black, Indigenous, People of Colour
[9] Posting personal information to incite violence
[10] Coordinating the sharing of damaging information across online platforms to give the appearance of

- Impersonation (63 percent); and
- Video- and image-based abuse (57 percent).

The violence stemming from these forms of abuse extends beyond the online environment. A recent New York Times investigation (Mozur, et al., 2024) identified a violence-promoting group hosted on the social media platform Telegram linked to a series of attacks, including a 2022 shooting at an LGBTQIA+ bar in central Europe. The phenomenon, whereby digital platforms facilitate the transition from online rhetoric to offline violence, is called *stochastic terrorism*.[11] It involves the use of mass communication to incite random individuals to commit statistically predictable but individually unpredictable violent acts. The content creators can often assert plausible deniability, claiming they did not directly incite violence. However, their actions contribute to an environment where such acts become more likely.

Case Illustration – Online mysogyny against female political leaders: Canadian example in the news

Threats, harassment and online hate driving women out of politics, MPs warn

Jasmeen Gill – The Canadian Press
March 8, 2025
Source: https://globalnews.ca/news/11073007/threats-harassment-and-online-hate-driving-women-out-of-politics-mps-warn/

Excerpt: **"**As longtime Liberal MP Pam Damoff prepares to leave politics when the next federal election is called, she is wistful but open about what is driving her to leave a career she has had for more than a decade. Vocal about the misogyny and threats she faced during her time in government, she wants public safety officials to take these threats more seriously. 'We've seen a shift in how people treat politicians, and I really worry that at some point, someone will be injured or killed,' Damoff said in an interview."

CBC News (Maimann, 2024) reported on recent research by various NGOs, that female politicians frequently face online abuse. This is attributed to "systemic social media problems" – particularly the lack of enforcement of community guidelines. Sobieraj (2020) highlighted rigorous research showing how online abuse of female officeholders is systematic and persistent. Female politicians and activists are often targeted with online threats, harassment and graphic sexual depictions - tactics designed to undermine their legitimacy, strip individuality, and discourage political engagement (DiMeco, 2019).

In April 2025, British news media reported on a UK parliamentarian who received a series of death and rape threats on social media, which she attributed to followers of social media influencers promoting misogyny (Barker-Singh, 2025). These online attacks began after she criticized a controversial social media owner. This case is noteworthy because it reflects a common feature in foreign information manipulation: certain channels align with actors

popular or grass-roots support

[11] https://en.wikipedia.org/wiki/Stochastic_terrorism

to support their objectives while evading attribution (e.g., Besancenot, 2025). In other words, these channels and followers act as unwitting proxies for other parties. This may also be considered a form of "soft violence", which refers to non-physical actions that, while not criminal, *per se*, are used to undermine social cohesion and assert group dominance, serving as a primary tool for communication, recruitment, and radicalization in violent transnational social movements (Kelshall, 2020).

In her critique of the polluted information ecosystem, Schick (2020) calls for a clear and consistent understanding of the disinformation problem as a crucial first step towards effective action. Heeding this call, we begin by framing online gendered disinformation as part of a broader context of harm and a set of harmful practices.

Conceptualizing Gendered Disinformation

Definitions

In 2022, the UN Women Expert Group sought to develop a common definition for the broad category of technology-facilitated violence against women or gender-based violence (TF-VAW/GBV). They determined that concepts of TF-VAW generally included some or most of the following features (UN Women Expert Group, 2022, pp. 3-4):

- **VAW or GBV:** An implicit reference to existing definitions of violence against women and gender-based violence;
- **Gender dimension/motivation of the act:** A specification that it is an act of gender-based violence, directed towards a woman because she is a woman or that affects women disproportionately. (We advocate for the inclusion of identities and behaviours that do not conform to traditional – particularly, ideologically-driven – formulations of gender within this dimension.)
- **Means:** Naming of ICT or technologies generally and/or specific technologies (e.g. spyware, GPS) as the means through which the violence was perpetrated.
- **Medium or Space:** Referenced as 'online' or 'cyber' or 'digital' spheres.
- **Forms of TF-VAW:** A list of some or several specific forms of TF-VAW, (e.g. sextortion, doxing, trolling).
- **Harm:** Reference to harms generally, or specific forms of harm, that ensue as a result of having experienced TF VAW (e.g physical, sexual, psychological, social, economic, other).
- **Continuum of VAW:** Reference to the fact that TF VAW occurs within a continuum of violence, that can include offline violence, and vice versa. For example, a woman may be stalked online and then the stalker may show up at her place of work, or a partner abusing a woman at home may monitor and control her movements event when they are not home, using GPS enabled technology.

As a result, the UN Women Expert Group (2022) proposed the following common definition for technology-facilitated violence against women, with the proviso that VAW could be replaced by GBV:

> ... any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms (UN Women Expert Group, 2022, p.4).

Within disinformation, a specific subgroup of TF-VAW – **gendered disinformation (GD)** – targets **women and girls *individually* through misogynistic harassment and intimate partner violence, *and collectively*** to subjugate them and deny them voice and participation in democratic society.

The National Democratic Institute defines GD as "the use of false information to confuse or mislead by manipulating gender as a social [wedge] to attack women and/or to sway political outcomes." (Jankowicz, et al., 2021, p.3). When spread online, GD may be viewed as a form of online violence, perpetuating hostile systems against women and posing "a credible threat to democracy" (Sobieraj, 2020, p.152).

Online (or digital) gendered disinformation involves the misuse of information communication technologies (ICT) to:

- Release/propagate false or misleading information about individual females, groups of females, or females, in general; *and/or*
- Overtly and explicitly abuse individual females, groups of females, or females, in general.

Case Illustration – Jess Davies and the long-term harms of online sexual exploitation

'I don't date at all now': one woman's journey into the darkest corners of the manosphere

When Jess Davies was 15, a boy leaked pictures she'd shared with him. At 18, she was a glamour model. A few years later, another man violated her trust. Then she fought back
Anna Moore – The Guardian
April 30, 2025
***Source*:** https://www.theguardian.com/society/2025/apr/30/i-dont-date-at-all-now-one-womans-journey-into-the-darkest-corners-of-the-manosphere?CMP=Share_iOSApp_Other.

Jess Davies, now a women's rights advocate and media professional, was first exposed to online sexual exploitation as a teenager when a private image she had shared with a trusted peer was circulated without her consent. Over the years, she became the target of repeated image-based abuse, including the unauthorized distribution of her photos, cyberflashing*, impersonation, and catfishing**. Her images were misused across pornographic platforms, social media, and anonymous forums, where users engaged in "games" that involved trading, modifying, and humiliating women through manipulated content and explicit commentary. In many cases, images were posted alongside the victim's name and contact information, encouraging coordinated harassment.

Davies' experience demonstrates how the distortion, misuse, or fabrication of content targeting individuals based on gender can intersect with sexual exploitation online. These harms were enabled by weak platform governance, societal stigma, and the absence of clear accountability for perpetrators. Despite years of digital abuse, Davies received no apology from those responsible. Her story also illustrates the long-term psychological, social, and professional impact of such violations, and the urgent need for legal reform, proactive platform responsibility, and survivor-centred support systems to address the growing threat of gendered disinformation and online sexual exploitation.

Case Illustration – Sharing intimate photos without consent: Canadian example in the news

In December, the Winnipeg Police Service were investigating reports of AI-generated nude photos of underage students circulating at Collège Béliveau, a Grade 9-12 high school in Windsor Park
Jen Zoratti – Winnipeg Free Press
February 10, 2024
Source: https://www.winnipegfreepress.com/arts-and-life/2024/02/10/seeing-is-believing-the-real-and-present-danger-of-fake-ai-images
Excerpt: "The speed and ease with which these images can be created and spread is also alarming; one doesn't even need to have a mastery of Photoshop anymore.... And yet, despite this rapid acceleration in technology, it seems as if we're still stuck in 2014 when it comes to the law. ... Manitoba is one of eight provinces that do indeed have intimate image laws, but ours don't refer to altered images. That needs to change, and fast. We cannot afford to have the creation and distribution of sexually explicit AI-generated images dealt with the same way online sexual harassment has traditionally been dealt with, which is to just tell women to 'stay off the internet.'"

Social Contexts and Conditions Enabling Gendered Disinformation

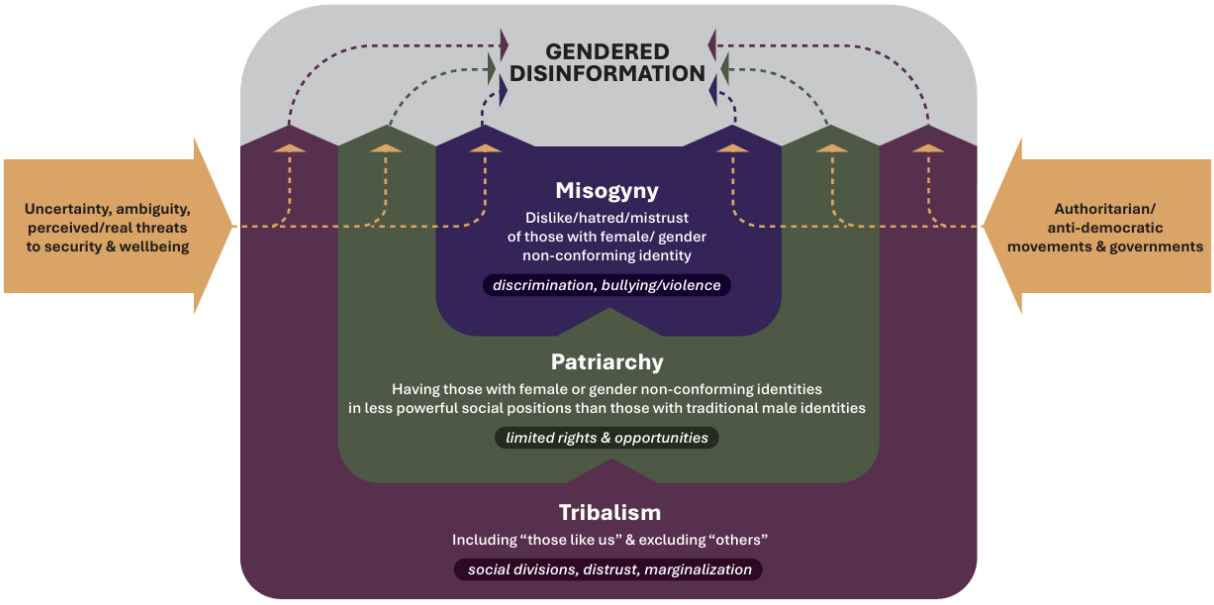
Gendered disinformation occurs within a complex context where cultural and social factors - such as group membership, identity, cognitive biases, beliefs, values, norms, and practices - influence human behavior and societal outcomes. This includes how interactions within cultural and social settings shape attitudes, actions, and development.

This context involves both socio-political and political-economic dimensions:

- How relationships among people, groups, and institutions – shaped by culture, group identity and social norms – influence political processes, policies, and power dynamics;
- How political institutions, processes, and policies – together with political decisions and the distribution of power and resources - affect economic systems and outcomes.

Key features of the broader social context that enable gendered disinformation, including online GD, include misogyny, patriarchy and tribalism (Figure 1).

Figure 1. Broader conditions may constitute fertile ground for the expression and spread of gendered disinformation.



Specific instances of GD may reflect one or multiple forms of systemic marginalization and injustice. These can include active repression – such as intimate partner violence and coercive control (e.g., Gill & Aspinall, 2020) – occurring domestically in what we term "intranational

repression" - or actions carried out at the behest of – or inspired by – foreign actors ("transnational repression") (e.g., Human Rights Watch, 2024).

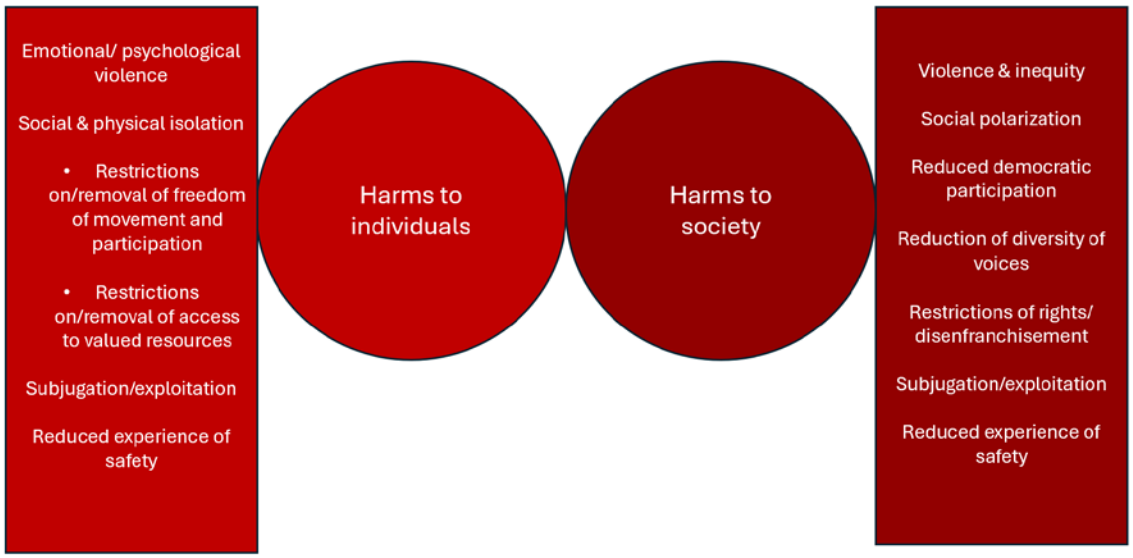
No single element of the broader societal context will, by itself, produce GD or its harms. These elements are akin to changing soil conditions, influencing how negative processes can be initiated, take root and impact their ecosystem. The socio-technical features of our world (such as digital technologies and political movements) create conditions that make either desirable or undesirable outcomes related to GD more or less likely. This means that *no single type of intervention can solve the problem*. However, understanding the layered contexts that can enable harm allow us to develop strategies to shift the environment toward more desirable outcomes.

- **Misogyny:** The hatred, dislike, or mistrust of those identified as being of the female gender. Manifests through discriminatory attitudes, behaviours, and institutional practices that demean, belittle, and oppress females, reinforcing patriarchal structures and limiting their social, economic, and political freedoms and opportunities (e.g., Sobieraj, 2020).
- **Patriarchy:** Male gender holds primary power and dominates in roles of leadership, authority, and control in both public and private spheres (e.g., Richardson-Self, 2021). This system often marginalizes females and limits their opportunities and rights.
- **Tribalism:** Inherited capacity for cooperation arising from cognitive & social practices – including the creation, propagation and promotion/enforcement of narratives – that reinforce shared identities & trust and which may intensify in-group/out-group dynamics (Samson, 2023).

In a mainly intra-national setting, the environment may include individually-focused gender-based violence and politically or ideologically-based harassment and abuse of individuals or groups. This may also involve political discourse or policy positions that reflect patriarchal or misogynistic rhetoric.

In a global setting, it may involve direct foreign interference in domestic life or democratic processes, or indirect foreign influence supporting misogynistic or patriarchal ideological movements (Figure 2).

Figure 2. Types of harms stemming from gendered disinformation.



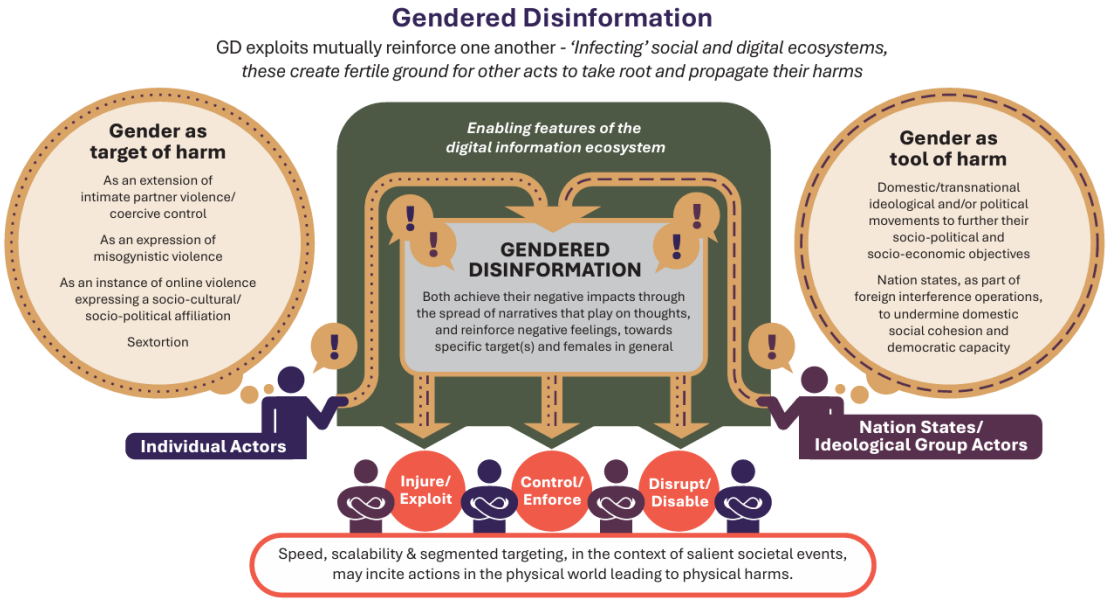
One desired effect is to deter women from participating in civic life and to erase from public policies values related to inclusion and belonging. Another is to exacerbate social divisions in the service of domestic political or ideological objectives, or adversarial geo-political goals.

These social and geographic contexts may also intersect, reflecting the influence of globalized narratives and movements. Perpetrators may be individuals, acting individually against females or groups of females. This positions female identity or gender as a focus of harm.

Group or nation-state perpetrators may target individuals perceived as obstacles to their global ambitions or ideological objectives (foreign interference against individuals), or as a method of disrupting the target nation's social harmony and democratic stability (foreign interference against the nation). These objectives may include the subjugation of women, girls and gender non-conforming persons, positioning female identity or gender as a tool for harm (Figure 3).

The broader socio-technical problems of contested truth and tribalism (of which populist movements are one example) have created a ripe environment for GD to proliferate. These exploits prey on individuals conditioned and motivated to believe false narratives and inaccurate explanations, including conspiracy theories centred on female identity or female/gender non-conforming leaders (e.g., van der Linden, 2023).

Figure 3. Position of females/gender non-conforming persons within gendered disinformation ecosystem.



Enabling Features of the Digital Ecosystem

Gender disinformation exploits are increasingly powered by AI advancements, such as botnets with natural language capabilities, and realistic synthetic media, ("deep fakes") (e.g., Schick, 2020). These emerging technologies complicate detection and countermeasures development.

AI-generated synthetic media, including deepfakes, voice cloning, and image-generation tools, allow the creation of convincing fake audio-visual content with minimal technical skill or cost (Lalonde, et al., 2025). These tools are widely accessible through user-friendly interfaces, democratizing the production of sophisticated disinformation.

The emergence of visual and multimodal disinformation (VMD) marks a significant shift in online abuse, including GD. Rapidly spread online, VMD technologies enable more persuasive, emotionally resonant, and harder-to-detect attacks (Lalonde, et al., 2025).

These capabilities have been weaponized to create non-consensual explicit content, falsely depict women – especially female politicians, activists, and journalists – in compromising scenarios, aiming to impersonate or discredit them. These tactics are not only invasive and damaging, but also serve to intimidate, silence, and discredit women in public life.

Members of racialized, LGBTQIA+ and intersectional communities are particularly vulnerable as potential "targets and tools", however this subgroup has been less well-researched (Thakur & Hankerson, 2021). Disinformation campaigns may draw on pre-existing, culturally potent,

discriminatory narratives related to both race and gender to lend credibility to false information. As a result, intersectional disinformation may weave together multiple harmful tropes and stereotypes, using these layered narratives to make false messages appear more believable and persuasive (Thakur & Hankerson, 2021). The impact of combining stereotypes with manipulated or manufactured audiovisual content, sometimes layered with actual news to enhance credibility, can be significant (Lalonde, et al., 2025).

These technologies, along with the design and business models underlying social media platforms, create opportunities for efficiency in scaling and targeting exploits of every kind. As Ressa (2022), Zuboff (2019) and others have demonstrated: features including the ease with which content may be re-posted within algorithmically optimized networks enables its speed and spread – its potential to amplify content faster than the pace of verification efforts (Lalonde, et al., 2025); the ease of access to networks which may be appropriated or botnets that can mimic popular support for a topic also contribute to the perceived realism of disinformation (Council of Canadian Academies, 2023); and, finally, the business incentive of platforms to attract and hold attention has led to the use of algorithms that arouse emotions and funnel increasingly intense content to users that have been caught-and-held (Bail, 2021). These "recommender" algorithms amplify certain voices and suppress others; they have also been flagged as a critical focus for mitigating online harms; there is a growing dialogue in the EU about regulatory options to disable or modify these algorithms (Ryan, 2025).

By leveraging these features of the digital ecosystem, online GD amplifies its disruptive and harmful effects while creating new avenues for victimization, both directly and indirectly. Indirectly, it fosters repressive elements of culture – hindering the safe and healthy participation in civic life. In some cases, these processes also create conditions conducive to stochastic terrorism, as previously identified.

COUNTERMEASURES: FOSTERING RESILIENCE AND DEVELOPING RESPONSE CAPACITY

The social scientific research base[12] available for developing GD countermeasures is just beginning to emerge. This has happened largely over the past several years, as part of efforts to better understand online identity-based polarization and abuse. Research thus far has varied, spanning

[12] This does not include military/national security research focused on the disruption of foreign interference and/or influence operations.

several disciplines[13] and drawing significantly from analyses of the mechanisms of mis- and disinformation. The focus of this work includes:

- Understanding the forms and mechanisms of disinformation and crafting appropriate interventions;
- Delineating the features and effects of online gendered violence;
- Critiquing the ways that the design, use and practices of social media platforms[14] foster polarization and harm;
- Analyzing the ways that populist disinformation produces and leverages social polarization as part of a broader set of socio-political objectives;
- Exploring and explaining how features of the digital ecosystem intersect with emotion- and identity-based factors to produce polarization; and
- Describing how certain contexts, together with desires for group affiliation and identity-protective cognitive processes, create vulnerabilities for disinformation threats that exploit narratives, symbols and other cultural artifacts.

False narratives[15] are used to harm or dehumanize members of designated out-groups by leveraging psychological biases among in-group members. These threats are designed to short-circuit critical thinking. Images and narratives that instil and amplify perceptions of scarcity and threats to the in-group's social standing blame out-groups for these issues.

Each of these focus areas provide information on the forms, mechanisms and impacts of information and narrative-based harms. They also identify ways to counteract disinformation, online harms, and to prevent or disrupt the social processes that create a fertile environment for online violence.

The Form of Disinformation Operations

Influence operations, including disinformation, often follow a structured format for achieving their objectives. This is known as a "kill chain". Drawing from military terminology, this term refers to a step-by-step outline of the stages of an attack – from initial reconnaissance to final impact. These frameworks help defenders understand and disrupt adversaries at each stage of their operation, rather than only responding to, or after, the attack.

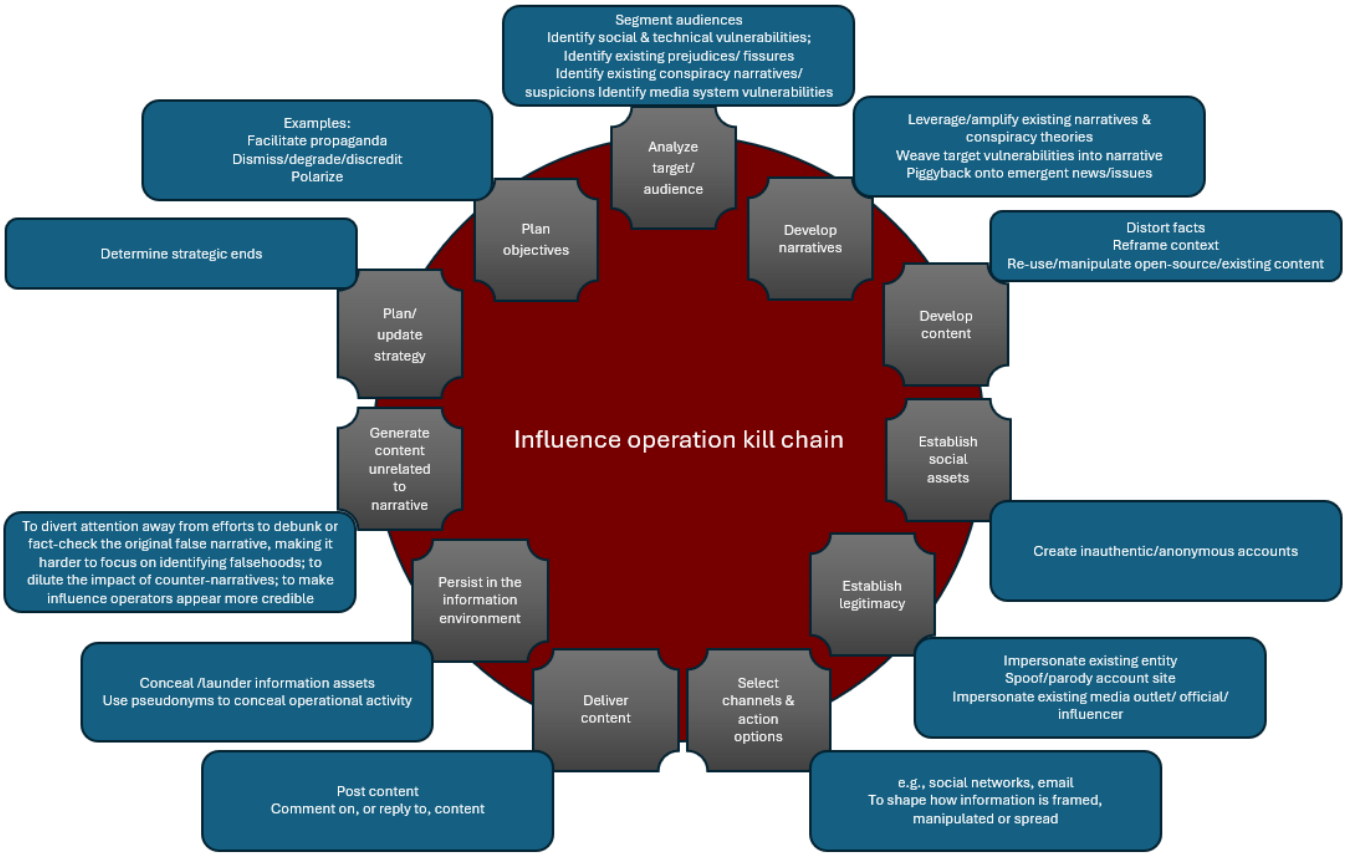
[13] Communications, sociology, social psychology, anthropology, as well as philosophy and history

[14] e.g., policies and practices on moderation and participation, business models, ownership issues, membership

[15] Content may include spoken or written narrative and/or images that invoke and align to particular narratives (e.g., evidence-free allegations that portray a woman as weak, unintelligent or incapable of leadership). As propaganda and influence operations have demonstrated, potent cultural symbolism (including music and other forms of artistic expression) may also be deployed as part-and-parcel of narrative interventions (see Pomerantsev, 2024).

The French Secrétariat général de la défense et de la sécurité nationale (VIGINUM)(2024) developed one such framework to provide insight into the operations used to carry out influence operations (Figure 4). This fairly intuitive structure can help us understand, identify and analyze the tactics, techniques and procedures that can be used as part of disinformation exploits, and determine practical avenues for responding to these where they occur.

Figure 4. General model of an influence operation kill chain (after VINIGUM, 2024).



The 11 components of the VINIGUM model cover areas such as: planning; development; delivery; and sustainability. It also suggests an iterative learning dimension to disinformation operations – based on assessments of responses to the exploit, operators may adjust various features of the campaign to refine targeting opportunities or to address to shifting goals.

However, kill chains are used largely in a responsive fashion, to analyze an attack once it has been detected, since these tools were designed for interdiction, not prevention. The presence of a problem should not be the starting point for security.

Spotting spreading disinformation

Example: The rapid spread of very similar-looking false information by numerous accounts in the immediate aftermath of an actual event – such as a political rally or a demonstration.

This might signal the use of a botnet or camouflaged account activity seeking to discredit or burnish the reputation of an individual or group. Appropriately packaged knowledge of the forms and features of disinformation operations can be an important part of awareness-building.

Understanding the typical format of disinformation operations can help targets, their supporters and responders detect, call out or interdict unfolding campaigns.

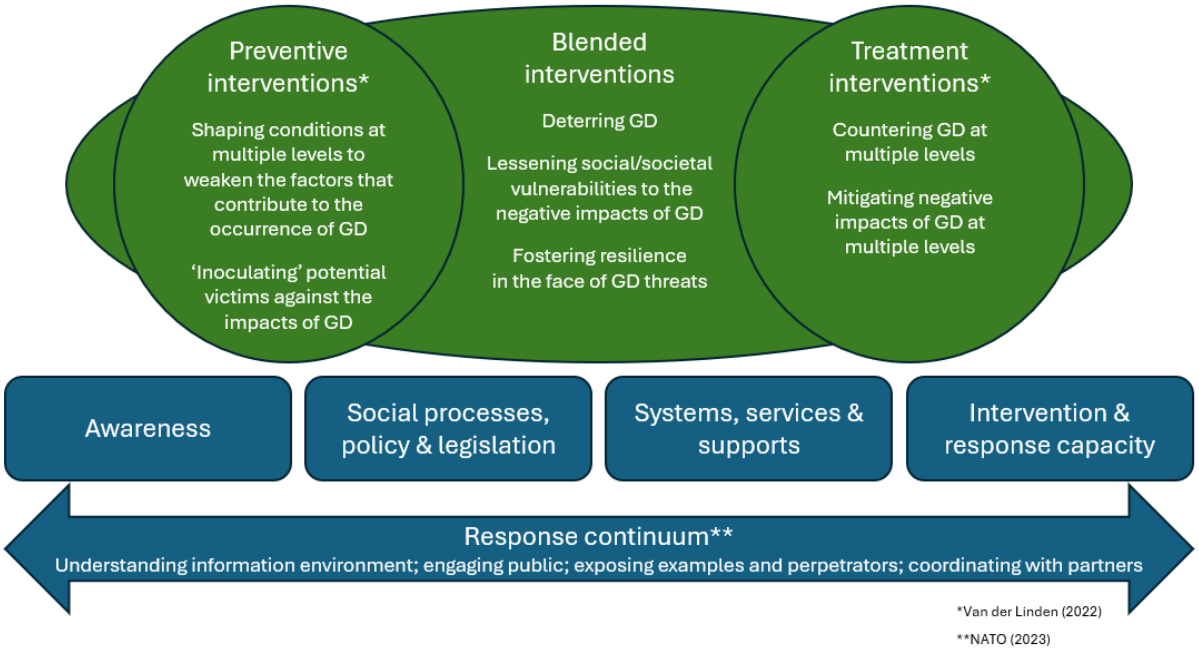
Strategic Interventions to Address Vulnerabilities and Threats

The EIU (2020) study of online violence against women identified that, at that time, efforts to tackle gender-based violence continued to focus mainly on post-experience responses, rather than on prevention. Despite this issue being flagged, the recommendations stemming from a recent report (Jankowicz, et al., 2024) predominantly focused on more 'downstream' or 'midstream' responses, such as platform accountability and victim support or responses to image-based sexual abuse.

van der Linden (2023) suggests that interventions can be helpfully conceptualized along a continuum of upstream to downstream actions categorized as largely preventive or largely "treatment" focused. Similarly, NATO (2023), recognizes that disinformation is simply one element of a wider collection of malicious information activities that range from hostile narratives targeting individuals to foreign information manipulation and interference (FIMI)[16]. At the far end of this continuum, hybrid warfare can make use of military and non-military channels to spread uncertainty among people and weaken the stability and trust within societies (NATO, 2023). Given the complexity of this landscape of digital harms, NATO's approach to countering disinformation emphasizes the importance of partnership and collaboration. Whether confronting GD as an expression of interpersonal violence, or as a geo-political phenomenon, a strategic mix of interventions, including those that fall "mid-stream" will likely be most effective as a framework for action – and as the basis for building greater resilience to the harmful effect is of GD. We propose an expansion of van der Linden's continuum to include mid-stream interventions at the macro and individual levels focusing on reducing vulnerabilities and strengthening resilience (Figure 5).

[16] While not always illegal, FIMI operations are manipulative activities designed to negatively impact "values, procedures and political processes in a target country" (NATO, 2023)

Figure 5. Proposed continuum of interventions for addressing gendered disinformation.



In only a few cases[17] – largely focused on mis- and disinformation not involving GD – has there been research on the effectiveness of specific countermeasures. Consequently, the development of countermeasures against GD, specifically, must be seen as an inferential exercise that ultimately should be tied to further research and evaluation.

An important caveat on any notion that tackling disinformation is a straightforward process is offered by Rid (2020), who observed that:

> Active measures have become more and more active and less measured to such a degree that they are themselves disintegrating – and this disintegration creates a new set of challenges. For the offender, campaigns have become harder to control, harder to contain, harder to steer, harder to manage, and harder to assess, For victims, disinformation campaigns have also become more difficult to manage, more difficult to assess the impact, and more difficult to counter.
>
> ...both open and closed societies... are both overstating and, more rarely, understating the threats and the potential of disinformation campaigns – and thus helping expand and escalate that very threat and its potential. (p. 434)

[17] Namely, work conducted by van der Linden (2023), Hameleers (2022) and Bail (2021) and their associates

This sobering assessment highlights the limitations of single-channel approaches to a complex, issue like GD. A broader vision for change is needed which focuses on shaping the conditions that increase positive outcomes and reduce negative ones.

Change of this magnitude must begin by mobilizing a diversity of experiences, perspectives and allies. It must include focus on individuals as well as groups and populations, and it must seek to counteract factors responsible for cognitive vulnerabilities (e.g., willful disbelief in facts[18]) and affective vulnerabilities (e.g., isolation, deprivation, status-seeking needs and perception of threats to social position[19]). A strategic vision of change should seek to:

- Enable awareness, identification and response capacity;
- Promote safety and pursue accountability;
- Foster a more equitable and inclusive society; and
- Support off-ramps to healthy identities and alternative affiliations for those vulnerable to enrolment in harmful movements and practices[20].

A crucial first step is to raise awareness and to make available concepts and systems for effective action by a variety of stakeholders. While individual action is important, GD is not simply a matter of "personal troubles"[21]. As Sobieraj (2020) and others have suggested, this is a public issue and a shared threat to democracies. Consequently, victims should not shoulder the burden alone – there is a role for everyone. While constructive changes against a problem of this magnitude will take time, systemic and coordinated change grounded in a holistic perspective, and informed by multiple voices, will be more effective than a series of unaligned incremental changes.

At the moment, the best available evidence on what may constitute promising countermeasures focus on points of vulnerability at which harms are either at risk of occurring, or have already begun to be associated with negative outcomes for people[22].

---

[18] e.g., McIntyre (2023), Norman (2021), Samson (2023)

[19] e.g., Bail (2021)

[20] As McIntyre (2023) suggests, in trying to wipe out the sources of the "disease" of untruths, we should also attend to the "sick" – i.e., those who have been deceived into believing and following GD narratives. However, this is difficult work, as van der Linden (2023) has described in relation to belief in conspiracy theories.

[21] Sobieraj (2020, p.138)

[22] Using a public health lens, and drawing from the Institute of Medicine's (IoM) framework of prevention (Pronk, Hernandez & Lawrence, 2013), preventive and blended interventions would correspond approximately to "universal" and "selective" measures, respectively, while treatment focused interventions would correspond to "indicated" measures. The IoM lens may have applicability to the problem of disinformation-polarization as it seeks to support effective action planning tied to an understanding of population-based levels of risks. Accordingly: universal prevention focuses on segments of the population deemed to be low-risk; selective prevention focuses on groups experiencing shared sets of risk factors (and may seek to boost protective factors); and indicated prevention seeks to serve those with emergent, detectable "signs and symptoms" of the problem of interest. The latter may involve individual and small-group delivery of services and supports aimed at preventing the progression of harms (Springer & Phillips,

Understanding and Awareness

Many have described the problem of disinformation as one of a "post-truth crisis" (e.g., McIntyre, 2023) or "infodemic" (van der Linden, 2022). This involves several features, including the creation of contested truths in relation to specific topics (such as climate). It also includes fostering more general conditions promoting contempt – even disgust – for those who have been positioned in some way as 'other', and cynicism towards the truths they share about their experiences (e.g., Bail, 2021; Samson, 2023). In the present case, 'others' are those members of female identifying gender groups that are positioned as objects of: subjugation; humiliation; shame; exclusion; blame; abuse; or any combinations of these harms. Is this light, gendered disinformation can be seen as a form of discourse-based violence.

Undermining the truth and the social position of designated 'others' serves an important aim of authoritarian ideologies within which misogyny is a central feature, and where those who identify as female are both tools and targets. The category of 'other' may include, for example, women, ethnocultural or religious minorities, members of political parties or members of the so-called 'elite' (media, academics, professionals, government and other public institutions, etc.). In some cases, there is overlap among these categories. Rid (2020), Ressa (2022) and McIntyre (2018, 2023) have shown how contested truth, the cultivation of cynicism and distrust, and the tools and platforms of social media, are used by populist authoritarians to undermine social cohesion and to target political adversaries by vilifying them through disinformation exploits that scale and repeat falsehoods. These are used to manipulate public opinion and to serve as a justification for persecution.

Blame and disbelief are key tools in the populist disinformation arsenal (Hameleers, 2022). Among other objectives, these are used to garner allies from among those who may lack healthy social connections and opportunities for secure and meaningful participation in society, and/or who fear the loss of valued, hierarchy-based, identities[23].

Contempt and Control

Sobieraj (2020) has described the ways that online gender-based attacks have the effect of creating a "context of contempt" and a "climate of unsafety" (p.35) that can undermine the willingness and capacity of women to participate in the everyday and democratic life of their communities and society. Corroding the social and political position of women is a key goal of misogynistic ideological movements and is known by its own term, Violence Against Women in Politics (VAW-P) (Jankowicz, Pepera & Middlehurst, 2021).

2021).

[23] These may include forms of supremacy that position specific groups at the top of a socio-political hierarchy which actively subjugates and blames those who are ascribed as being outside of this category.

As Richardson-Self (2021) has observed, recurring discourse focusing on subordinating identity-based groups is often accompanied by a "constellation of other acts" (p. 81), which can include forms of involuntary control and actual or threatened physical violence (Havard & Lefevre, 2023). The range of these tactics, which are used to enforce male dominance, can be referred to as "coercive control" (Stark, 2007).

> *Democracy requires a common understanding of reality, a shared view of what has happened, that informs ordinary citizens' decisions about what should happen, now and in the future. Authoritarians target this shared understanding, seeking to separate us from our own history to destroy our self-understanding and leave us unmoored, resentful, and confused. By setting us against each other, authoritarians represent themselves as the sole solution.*
>
> (Stanley, 2024)

Coercive control involves both physical and non-physical tactics to dominate and isolate the victim (Gill & Aspinall, 2007, 2020). These tactics include threats, monitoring, financial control, and restricting access to loved ones, ultimately eroding the victim's sense of self and freedom (Gill & Aspinall, 2020; Stark, 2007). It is considered an infringement on basic liberty and a form of intimate partner violence when it occurs within these types of relationships – whatever the gender or sexual orientation of the parties.

Common aspects like controlling actions, psychological abuse, sexual jealousy, and stalking can be facilitated by information communication technology (ICT) (Dawson et al., 2019). Douglas, Harris, and Dragiewicz (2019) found that ICT tools, such as smartphones and IoT devices, are used for technology-facilitated violence.

Victim-blaming and disbelief, often seen in coercive relationships and failed responses, align with misogynistic and populist ideologies (Bail, 2021; Cuklanz, 2023; Richardson-Self, 2021; Samson, 2023; Sobieraj, 2020). Similarly, in cyber deception, tactics like uncertainty and misdirection are key (McMahon, 2021).

Foreign Inteference and Manipulation of Information

Like all forms of disinformation, gendered disinformation interferes with the capacity of a society to engage in constructive public dialogue involving a pursuit of common understanding based on shared facts (Richardson-Self, 2021). As a result, gendered disinformation has been used as a component of FIMI operations – efforts by foreign governments or actors to influence public opinion, political decisions, or social stability in another country. This is done by spreading false or misleading information, often through social media, news outlets, or other communication channels. These activities are usually designed to create confusion, distrust, or division among people, and they can target elections, public health responses, or social issues. FIMI is considered a serious threat because it can quietly undermine a country's democracy, security, and public confidence without using traditional weapons or direct attacks.

Work by Bradshaw and Henle (2021) explored how foreign state actors (Russia, Iran and Venezuela) conducted covert influence operations on the Twitter platform to target Western feminist activists and politicians.

Several strategies were used by state-associated operatives to undermine feminist advocates/feminist narratives indirectly or directly. These included narratives designed to:

- **Promote in-group solidarity and out-group divisions** (e.g. around racial and political identities) to amplify negative feelings towards a movement and its supporters – such as that activists were "man-hating" and oppressive;
- **Undermine women's shared sense of a collective identity** by co-opting internal critiques within feminist movements; and
- **Direct online harassment and character attacks against individuals** to delegitimize or discredit them – in some cases, these were combined with threats of physical violence.

The latter were found to be more likely to occur via direct messaging to victims, rather than on public platforms (possibly because Twitter, at that time, was more actively deploying automatic detection measures). This research highlights how digital interference operations involving techniques for promoting social divisions and disrupting collective action are being used to undermine gender equality and weaken democracy by making it harder for women to speak out and mobilize for change (Bradshaw & Henle, 2021).

A related application is gender-based transnational digital repression. This involves the use of TF-VAW by authoritarian regimes to interfere with the exercise of free speech and activism by female diaspora residents of other countries. Research by the Citizen Lab at the Munk School of Global Affairs and Public Policy has shown how invasive monitoring and other forms of surveillance, along with online harassment and various forms of reputational assaults, have been used to extend the control of distant repressive states, or to marginalize victims within their diaspora communities in Canada (Aljizawi, et al., 2024; Michaelsen & Anstis, 2025).

For example, attackers have used mercenary spyware[24] implanted on devices using phishing exploits[25] to collect information and monitor the activities of civil society targets. The spyware is

[24] For example, tools developed by the NSO Group, as described by Deibert (2025), who provides detailed accounts of the use of mercenary spyware against civil society actors.

[25] Phishing exploits are deceptive tactics used by cyber operatives to trick individuals into revealing sensitive information, such as passwords, credit card numbers, or banking credentials. These schemes often involve fraudulent emails, text messages, or websites that closely mimic legitimate sources – like a bank or trusted organization – in order to gain the victim's trust and steal personal data. For example, an email may appear to come from an individual's bank asking them to "verify your account," when, actually, it is a carefully crafted exploit.

used as part of initial reconnaissance activities designed to determine social and technical vulnerabilities as part of the prelude to a disinformation operation. These kinds of exploits follow carefully designed, nested, operational plans, drawing from features of the kill chain used to implement malicious cyber exploits in the context of format of influence operations (Figure 6).

Cyber-enabled exploits are an important feature of contemporary influence operations (including information warfare) because they provide sophisticated hostile actors (individuals, groups or governments) with a number of benefits, such as:

- Securing new sources of private information;
- Diverting attention from the main objectives of an information operations; and
- Interfering with counter disinformation capabilities (Whyte, 2020).

Spyware – a powerful new threat

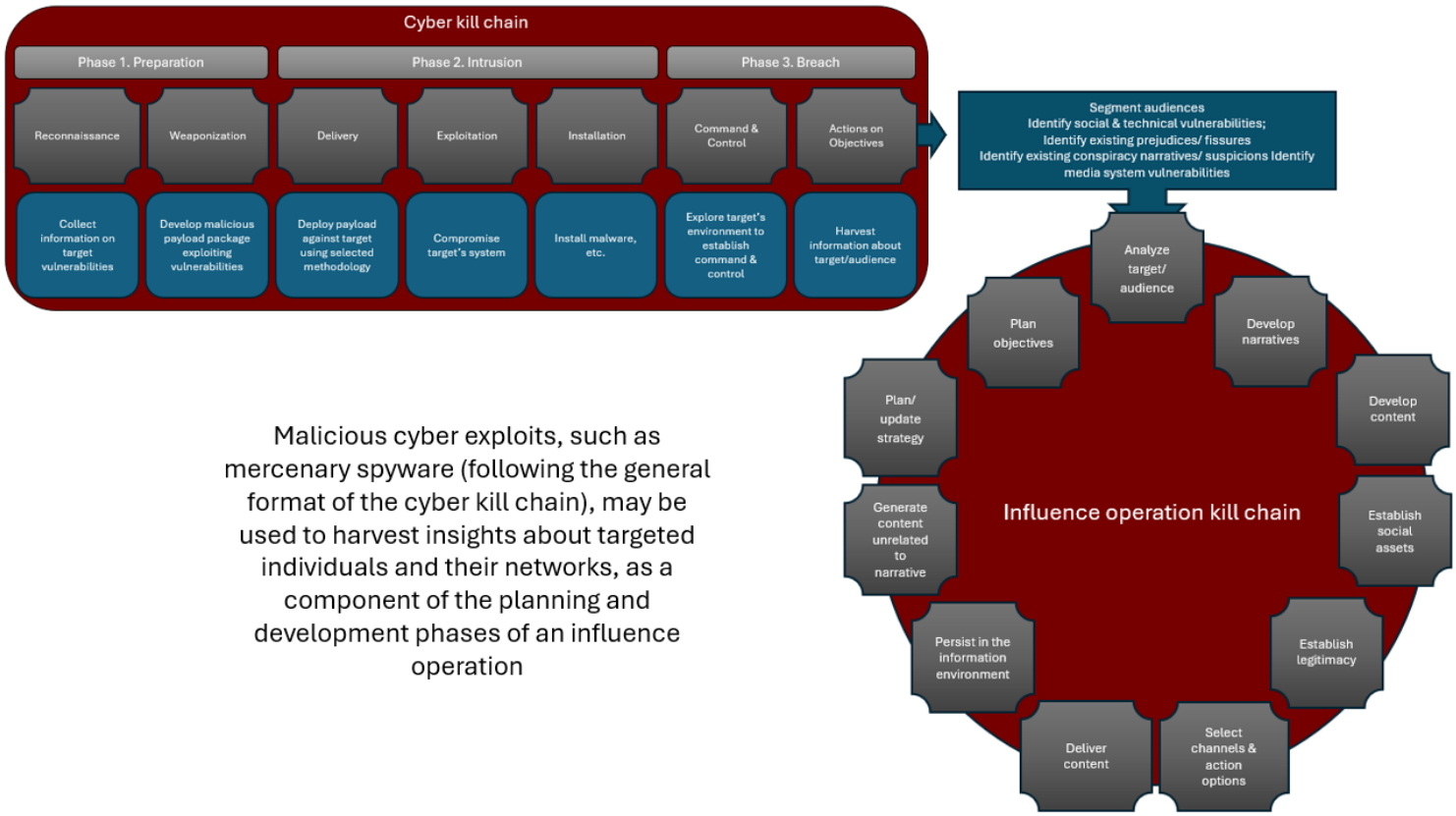
Malware or spyware exploits are tools used by attackers to secretly access or control computers and devices. In influence operations, these malicious programs can be used to steal sensitive or personal information, or monitor activities, as insights used to craft a disinformation campaign. These tools can also be used to manipulate communications, or spread false narratives to achieve political, social, or economic goals.

Imagine someone receives a fake email that looks like it came from their social media site. When they click the link, malware is installed on their device. The attacker uses this malware to steal login details, take control of the account, and then spread disinformation or harmful messages to all the person's contacts, making it look like those messages came from someone they trust.

Michaelsen and Anstis (2025) observed that many of today's authoritarian regimes see traditional gender roles and patriarchal norms as enablers of the social hierarchies and agendas on which the regimes depend for their bases of power and control. They assessed that these structures of domination need to be sustained both domestically and internationally in order to safeguard the stability of the regime. Where misogyny is used as an enforcement tool, polarizing potential adversarial alliances and disrupting in-group cohesion among their opponents helps authoritarian leaders legitimize their agendas in the eyes of their ideologically aligned supporters (Michaelsen & Anstis, 2025).

Among its many uses, spyware is being deployed by authoritarian governments as a tool for supressing dissent. An example of the use of cyber-enabled repression was uncovered by Citizen Lab (Marczak, et al., 2021) which described how two types of mercenary spyware – Cytrox's Predator and NSO Group's Pegasus – were used to compromise the iPhones of an Egyptian journalist and a politician via WhatsApp messages. In one case, both of these forms of spyware were used against the same individual. These compromises were carried out as part of an operation designed to quash dissident voices within civil society.

Figure 6. Example of the way that elements of a typical cyber exploit may connect to those of a targeted influence operation (after Hutchins, Cloppert & Amin, 2010 and VINIGUM, 2024, respectively)



Malicious cyber exploits, such as mercenary spyware (following the general format of the cyber kill chain), may be used to harvest insights about targeted individuals and their networks, as a component of the planning and development phases of an influence operation

Psychological Vulnerabilities Exploited by Disinformation

Education and awareness are important elements of an effective response at the societal level as well as in relation to opportunities for accountability, redress and rehabilitation by perpetrators (e.g., McIntyre, 2023; Cuklanz, 2023).

Understanding how online repression works and, in some cases, using digital tools to uncover and address online abuse are key to combatting gendered violence (e.g., Parrish, et al., 2023; Carty, 2023; Cuklanz, 2023).

One of the ways that fake content works is by showing us things that we anticipate would go together. The truth-likeness of these texts, images, or sounds can fool our systems of perceiving and thinking clearly by capitalizing on everyday expectations, otherwise known as cognitive biases.

Cognitive Biases

Cognitive biases are commonly understood as systematic and widespread mental tendencies that distort how we process information, often leading to outcomes that are inaccurate, flawed, or less than optimal (Korteling & Toet, 2020). They are mental "shortcuts" or patterns of thinking that can lead people to make judgments or decisions that are not fully rational or accurate. They also make us vulnerable to manipulation (McIntyre, 2018).

> ***Cognitive biases are hard to notice***
>
> *Succumbing to cognitive bias can feel a lot like thinking. But especially when we are emotionally invested in a subject, all of the experimental evidence shows that our ability to reason well will probably be affected.*
>
> (McIntyre, 2018)

These biases arise because the brain tries to simplify complex information or make quick decisions under uncertainty. While these shortcuts can be helpful in daily life, they can also cause people to misinterpret information, overlook evidence, or rely too heavily on pre-existing beliefs.

Cognitive biases are one contributor to the believability of disinformation (McIntyre, 2018). Previous research conducted by George, et al. (2021, cited in French, et al., 2025) concluded that confirmation bias – the tendency to favor information that aligns with existing beliefs and dismiss information that contradicts them – was influential in the spread of fake news. McIntyre (2018) suggested that other cognitive biases may be notable for their contributions to a vulnerability to believing false information. These include:

- **Backfire effect:** In which people strengthen their existing beliefs when presented with information or evidence that contradicts them. Instead of changing their views, they double down, often becoming even more committed to their original position. This effect highlights one reason why individuals may resist changing their beliefs, even in the face of clear, corrective information.
- **Dunning-Kruger effect:** The tendency for individuals with limited knowledge in a subject to overestimate their understanding, making them more susceptible to disinformation. This is a version of what is also known as overconfidence bias (see below).

In a recent study, French, et al. (2025) examined social media users' perceptions of how they use and share fake news. They analyzed these results to identify the cognitive biases that appear to shape the believability of fake news when it is being consumed. They found five cognitive biases likely to make fake news believable. These were:

- **Herd mentality:** The tendency of individuals to adopt the thoughts or behaviors of a group, often following peers rather than forming independent judgments. This dynamic can lead people to accept information as true simply because it aligns with the majority view, rather than critically evaluating the evidence themselves.

- **Confirmation bias:** The tendency to favor information that aligns with existing beliefs and dismiss information that contradicts them.
- **Framing cognitive bias:** Involves the ways that people's decisions are influenced by how information is presented, rather than by the facts themselves. Even when the underlying information is identical, different wording or context can lead to different conclusions. In the context of fake news, this bias can cause individuals to judge a story's truth based on how coherent or compelling it sounds, without verifying the information.
- **Overconfidence bias:** This involves overestimating the accuracy or depth of one's knowledge. This can lead people to have excessive confidence in their judgments or decisions, even when their actual understanding is limited.
- **Anchoring bias:** The reliance on the first piece of information encountered, which can shape how new information is interpreted, even if the first source is false.

These biases can cloud judgment and make disinformation seem more believable or harder to challenge. When content aligns with a number of these biases and is congruent with the context in which it is being experienced, it can be difficult to see it as false (van der Linden, 2023).

Despite these difficulties, French, et al. (2025) proposed a set of seven empirically-informed measures that have promise for mitigating the risks posed by these biases. Most of these methods address more than one cognitive bias.

- **Consider-the-opposite strategy:** Encouraging users to actively consider opposing viewpoints—such as linking to articles with contrary perspectives. This can help reduce anchoring bias by broadening the information considered when evaluating truth claims.
- **Analysis of Competing Hypotheses (ACH):** Presenting multiple, competing explanations for a narrative prompts users to evaluate corresponding evidence rather than relying on how information is framed, thereby mitigating confirmation and framing biases.
- **Opt-in obfuscation:** Requiring users to actively choose to view potentially biased or misleading content, by clicking through a warning, can disrupt automatic processing and reduce confirmation and anchoring biases, encouraging exposure to attitude-opposing information.
- **Dynamic flags:** Unlike static warnings, dynamic flags (e.g., blinking alerts or overlays) that require user interaction have been shown to better attract attention and mitigate overconfidence bias, especially when combined with other cues like ACH or evidence ratings.
- **Evidence rating:** Asking users to rate the credibility of content (without enabling content removal) shifts their focus toward evaluating information on its merits. This can reduce framing bias, particularly when users are prompted to think critically about source trustworthiness.

- **Text visualization:** Using visual tools like word clouds disrupts the linear reading of text and reduces the influence of emotionally charged framing. This can help mitigate framing bias by refocusing attention on content rather than narrative style.
- **Hide virality statistics:** Removing public engagement metrics (likes, shares, views) from news content can reduce herd mentality bias by preventing users from using popularity as a proxy for truth.

All of the preceding methods lend themselves to alterations in the user experience design of news and social media platforms. The consider-the-opposite strategy and ACH also suggest opportunities for awareness training and the development of personal reflective habits that can help consumers of online content engage with these platforms with a greater degree of agency. However, some of the core features of disinformation are difficult to resist. This is particularly the case when content is encountered repeatedly.

Repetition is "Sticky" and Contagious: The Truth Illusory Effect and Message Virality

van der Linden (2023) describes a set of studies that demonstrated that, the more a message is repeated, the more true it feels. This is known as the illusory truth effect[26]. Research has shown that even when there is prior knowledge of a particular topic, this does not by itself protect against false truths (van der Linden, 2023; Fazio, et al., 2015). This is more the case when the content is being echoed by what are seen to be credible sources (McIntyre, 2023), such as media outlets that amplify messages without accompanying critical analysis.

When a message is repeated with a high frequency, delivered with fluency, and/or experienced as emotionally charged, and/or received in the midst of felt pressure, it may be harder still to detect as false[27]. For example, the viral nature of conspiracy theories, and their psychological potency – what van der Linden (2023, p. 49) terms an "evidence-resistant worldview" – make them particularly dangerous. He (2015, 2023) reports that even brief exposure to a conspiracy theory can render people less civic-minded. Moreover, a disposition towards a conspiratorial worldview[28] tends to result in people taking one conspiracy theory as evidence of others, however implausible any of these may be (Biddlestone, Azevedo & van der Linden, 2022; van der Linden, 2015, 2023).

The more the message becomes familiar, and the fewer the opportunities for critical reflection, the more likely it is that the message will be perceived as true. Not only might this influence the thoughts and actions of individual information consumers, to the extent that these individuals experience the information as true, they will participate in its amplification by sharing it casually or intentionally among members of their own networks (e.g., McIntyre, 2023; Ressa, 2022).

---

[26] Hasher, Goldstein & Toppino (1977); Fazio, Brashier, Payne & Marsh (2015); Fazio & Sherry (2020)

[27] An important and powerfully negative antecedent of this insight was Hitler's 'Big Lie Rule' of propaganda which involved the assertion that if a big enough lie is told often enough, most people will come to believe it (for additional detail, see van der Linden, 2023 and Pomerantsev, 2023).

[28] Also know as a "monological belief system" (van de Linden, p. 49)

However, knowledge of disinformation – what it looks like and how it works – is thought to be protective (van der Linden, 2023; McIntyre, 2023; Ressa, 2022). A body of research by van der Linden and colleagues summarized in van der Linden (2020) identifies a consistent set of seven features of conspiracy theories that can be used as a short-hand to identify this form of mis- or disinformation. These are summarized by the mnemonic, **CONSPIRE**:

- **Contradictory:** The narrative contains internal contradictions – it doesn't "hang together" logically;
- **Overriding suspicion:** The narrative expresses suspicion about any official positions about the topic;
- **Nefarious intent:** Sinister intentions are attributed to those who are thought to be the conspirators;
- **Something must be wrong**: Believers might let go of aspects of the story but still insist that "something must be wrong";
- **Persecuted victim**: Believers often view themselves as victims of plots created by powerful elites;
- **Immunity to evidence**: Challenges to the conspiracy story are interpreted as evidence of the conspiracy; and
- **Re-interpreting randomness**: Random events that don't seem to have anything to do with the conspiracy story are interpreted as evidence for the conspiracy, even though another cause of the event is more likely.

Spreading knowledge of the typical format of conspiracy theories, and encouraging practice in using this knowledge to notice and analyze false narratives, may help people build resistance to the insidious effects of repetition.

The Role of Identity and Affiliation Needs

Research on tribalism (Samson, 2023) and affective polarization (Bail, 2021) shows how disinformation takes advantage of strong emotions and people's needs related to identity and belonging to make it harder for society to stay united or agree on shared facts.

One of these levers of manipulation is the human tendency to be attracted to social contexts that preserve or build up self-worth among those who feel that their preferred identities and social position are under threat (Bail, 2021; Pomerantsev, 2023).

The "manosphere" has gained attention in recent years. In a 2014 Washington Post article on the perpetrator of misogynist terror attacks near the University of California, Santa Barbara[29], journalist

[29] Wikipedia (n.d.). 2014 Isla Vista killings. https://en.wikipedia.org/wiki/2014_Isla_Vista_killings

Caitlin Dewey described the it as: "that corner of the Internet where boys will be boys, girls will be objects, and critics will be 'feminists,' 'misandrists' or 'enemies.'" It is a vast network of blogs and forums that promote hyper-masculine ideologies and hostility toward women and feminism. While not all components are violent, the core belief is that feminism has corrupted culture and that men should reclaim dominance by embracing traditional gender roles.

A more recent data-driven investigation of these online spaces by Ribeiro, et al. (2021), based on a taxonomy first developed by Lilly (2016), characterized it as a growing and prospering "conglomerate of web-based misogynist movements focused on 'men's issues'" (p. 196). A core shared belief across these online communities is that "masculinity is under siege by feminizing forces; and feminism is hypocritical and oppressive" (Ribeiro, et al., 2021, p.197). Using Lilly's (2016) taxonomy, Ribeiro, et al. (2021, p.196) described four prominent communities within the manosphere:

- **Men's Rights Activities (MRA):** Advocate for men's issues, arguing that social institutions unfairly disadvantage men. This movement is often characterized as misogynistic.
- **Men Going Their Own Way (MGTOW):** Promote the rejection of relationships with women and mainstream society, rooted in the belief that the system is irredeemably biased against men.
- **Pick Up Artists (PUA):** Teaches men manipulative techniques to attract women, frequently involving objectification, harassment, and a belief that modern masculinity is undermined by female dominance.
- **Involuntary Celibates (Incels):** Mostly young men who bond over feelings of sexual rejection and resentment toward women, often expressing violent or self-destructive ideologies linked to real-world acts of violence.

Examining data across a 14-year period, Ribeiro, et al., (2021) found that older sub-groups (MRA, PUA) had declined in popularity and activity, while newer, more extreme – "toxic" – sub-groups (MGTOW, Incels) were "thriving". They concluded that the manosphere is evolving from what was previously a looser conglomerate of related communities towards a cohesive whole, where people are participating in more than one sub-group. This environment also seems to be fertile ground for the emergence and growth of more extreme sub-communities, connected by their adherence to "Red Pill"[30] ideas (Ribeiro, et al., 2021).

[30] Beliefs, often shared in online communities, claiming to expose hidden truths about society, gender, and power, often in opposition to mainstream values. Borrowed from the 1999 film *The Matrix*, the term originally symbolized awakening to reality. Online it is often linked to misogynistic, anti-feminist, and male supremacist ideologies. In these spaces, "taking the red pill" means rejecting feminism, believing men are oppressed, and embracing rigid gender roles. Common in the manosphere (e.g., MRAs, MGTOW, Incels), red pill rhetoric is often as a gateway to extremist content. Ribeiro (2021) found that, by the end of 2018, the */r/TheRedPill* subreddit ranked third in total posts and fourth in monthly active accounts.

Combined with the addictive design of social media platforms (e.g., Lanier, 2018; Zuboff, 2019), online environments that appear credibly to validate and intensify feelings of injustice and outrage create a ripe environment for disinformation. For example, video games have begun to receive attention for the potential role they may play in exposing young men to radicalizing online communities (e.g., Sorell & Kelsall, 2025; Stuart, 2025).

Where individuals are vulnerable to being influenced by the threat of further perceived losses, or additional social exclusion, being simultaneously being welcomed into a fraternity of fellow 'victims', can be a powerful experience (Bail, 2021; Samson, 2023).

Maté (2022; 2024) described in detail the ways that prior histories of trauma, including harsh or abusive early years, may neurologically predispose certain individuals to the malign influences of radicalization. In particular, Maté suggests that experiences of severe trauma lie at the root of risks for enrolment in extreme authoritarian movements. In addition to offering a refuge from feelings of vulnerability, these movements also invite a sense of belonging for those who harbour grievances related to perceived or real experiences of exclusion, dislocation or marginalization (Maté, 2024).

Echoing the role of context in how online messages are received and interpreted, Bail (2021) argues that what he termed the "social media prism" both reflects the broader social landscape back to users, and distorts what is being seen in ways that may create an altered and misguided form of self-worth. He describes how this distorting, but perversely empowering, experience makes it easier to carry out extreme online actions for those who regularly experience dis-empowerment in their off-line lives.

These destructive online behaviours are ways to signal membership in alternative identity-affirming groups. Samson (2023) suggested that identity-protective cognitive processes play a significant role in shoring up disbelief in truth and belief in conspiracy theories. Bail (2021) proposes two relational processes that scaffold increasingly extreme online behavior: the normalization of extremism as a taken-for-granted feature of one's reference group (and a misapprehension that one's reference group is more of the norm than the exception); and an exaggeration of the extremism of opposing sides. Bail argues that these processes, which make a person's own extremism appear reasonable and that of others seem more extreme, creates a feedback loop that intensify extreme thoughts, feelings and behaviour. Where these loops also shore up identity protective processes, and symbolize membership in status affirming groups, they are powerful barriers to change.

Norman (2021) sought to examine ways that toxic ideologies could take hold to inspire a range of harms based on the notion that "bad ideas are mind parasites" with infectious properties (p. 3). Drawing from a range of research and theory, he suggested that, just as physical stress could weaken the body's immune capacity, so could psychological and cultural stress weaken the mental immunity of individuals and groups to cognitive 'pathogens', such as divisive ideologies. He suggests that, not unlike the ways that human viruses propagate by infecting one person and then another, harmful ideologies are spread from person to person and can be amplified through

technologies. They find fertile ground where there exists a tendency towards belief and an active rejection of invitations to disbelief.

Norman (2021) and Samson (2023) suggest that 'mental immunity' can be developed by cultivating people's innate capacity to detect, filtre and remove bad ideas. It is anchored in a practiced capacity to: evaluate ideas presented to us; an openness to constructive doubt about what we are seeing or being told; and a willingness to revise our opinions. Both Norman and Samson suggest that, when individuals, or cultures, fail to nurture these capacities, the result can be a context that promotes, rather than inhibits, harmful ideas. Samson (2023) identifies the rejection of openness to doubt – or willful belief – as a critical vulnerability that disrupts the "linkage between critical thinking and belief revision" (p. 345). Where a key symbol of membership in an identity-affirming group is "willful unreason" (Samson, 2023, p. 345), group members may be more vulnerable to disinformation that is consistent with group belief systems and more resistant to narratives that invite reasoned alternative accounts.

Psychological Propensity to Ideological "Capture"

A new area of research has begun to explore a significant vulnerability to enrolment in disinformation about gender: a psychological propensity towards ideological thinking. Rather than focusing only on the content of belief systems, this work explores the cognitive bases of ideological thinking itself, suggesting how belief formation and susceptibility to disinformation and conspiracies may interact.

Zmigrod and colleagues (Zmigrod, 2022; Zmigrod, et al., 2023) describe an ideological style as being marked by rigid adherence to doctrine, resistance to updating beliefs in the face of new evidence, and strong loyalty to in-groups, often coupled with hostility toward out-groups. They propose that ideological thinking—regardless of its specific content (e.g., political, religious, or conspiratorial) – shares a common psychological structure. This research suggests those who have a strong need for certainty and a low tolerance for ambiguity, for example, are more likely to seek out belief systems that offer clear-cut answers and structured explanations of the world. Psychological, social-emotional and situational factors appear to play a role.

At the psychological level, these individuals appear more likely to exhibit cognitive rigidity – difficulty updating beliefs when presented with new evidence. They may also show a need for cognitive closure – preferring firm conclusions over uncertainty, which can make them especially receptive to dogmatic ideologies that promise order and clarity.

On the social and emotional level, ideological thinking is often reinforced by experiences that promote strong in-group identity and a sense of belonging. Recalling work done by Bail (2021) and Sampson (2023), individuals who feel a deep emotional connection to a group – whether political, religious, or cultural – may be more likely to embrace ideologies that emphasize loyalty and divide the world into "us" versus "them." Zimgrod and colleagues found that this can be accompanied by hostility or distrust toward out-groups, which further entrenches belief systems and resistance to

alternative perspectives. Additionally, individuals with a high need to belong may gravitate toward ideologies that offer not just explanations, but also community, purpose, and meaning.

Situational factors also play a role. In times of instability – such as economic hardship, social unrest, or high levels of misinformation – people may cling to rigid belief systems as a coping mechanism (Zmigrod, et al., 2023). Ambiguous environments, like social media, can further amplify these tendencies by encouraging quick, emotionally driven responses over critical reflection. Taken together, these findings help explain why some people are more vulnerable to ideologically-charged disinformation, including gendered disinformation exploiting identity, fear, and emotional resonance.

As non-egalitarian gender belief systems tend to be grounded in rigid, binary gender roles (presumed to arise from inherent biological differences), these findings map onto factors fostering gendered disinformation, including tribalism, patriarchy and misogyny, and the role that mis- and disinformation and conspiracy theories play in propagating negative gender-based narratives.

The idea of a common psychological propensity towards the adoption of extreme or rigid belief systems, may offer insights into individual susceptibility to disinformation campaigns targeting gender or identity. Because, these ideologies frequently involve hostile or discriminatory treatment of those who deviate from these normative expectations, Zmigrod's research may provide a window into understanding, and intervening in the face of, risks that could escalate to more serious harms.

Another facet of this research is the finding that when individuals are assessing the reliability of incoming information against their prior beliefs, "noisy" or uncertain information environments – like social media can skew this process, making people more likely to accept false information, especially when it aligns with their existing ideological worldviews.

At a broader, societal level, the concept of "rape culture" has increasingly been used to explain how sexual violence is normalized and accepted within digital spaces (Sugiura & Smith, 2020). This social permission structure and social learning environment encompasses a wide range of gendered norms, behaviors, attitudes, beliefs, values, customs, symbols, language, and practices that tolerate, excuse, or even promote or celebrate sexual aggression (Powell and Sugiura, 2018, cited in Sugiura & Smith, 2020). Sugiura and Smith (2020) suggest that, while the concept originally focused on cis-gender[31] women's experiences in heterosexual contexts, it also provides a valuable framework for understanding the power imbalances underlying sexual violence, abuse, and harassment targeting LGBTQIA+ individuals.

These studies suggest that the way people respond to disinformation is shaped not just by political views or media literacy, but by deeper psychological processes or structures, involving elements of ideological rigidity and in-group identity attachments. These widely shared features may have held

[31] The term, cis-gender, refers to a person whose gender identity matches the sex they were assigned at birth.

evolutionary benefits. Today, individuals susceptible to ideological thinking may be particularly vulnerable to gendered disinformation, which often leverages emotionally charged, identity-based narratives. While the research does not focus on gendered disinformation specifically, it offers insights into who may be most at-risk and underscores the need for future work to explore how cognitive style and trust in information shape susceptibility. These insights could help inform more targeted and effective interventions.

Raising awareness of gendered disinformation (GD) helps people see it as a shared public problem. This understanding is crucial for prevention efforts (Jankowicz et al., 2024). It not only mobilizes attention and action but also encourages public discussion, also involving youth, about possible solutions. From a prevention standpoint, these interventions may work to reverse the normalization of these online practices.

Implications for Countermeasures

Fake truths become more real and more 'sticky' the more they are repeated and amplified by credible sources, especially within a context that is consistent with the content of the message. This is the case, even when the targets of GD have prior knowledge about a topic.

Illusory truth is a particularly powerful vector of disinformation. However, its perceived truth-likeness can be reduced when target audience members are aware of disinformation in general, and the falsity of an individual message (or source) in particular.

Abusive narratives that are embdedded within conspiracy theories which have been amplified across social media platforms may be resistent to redress. This is because un-identified conspiracy theories have a high degree of believability – particularly in certain contexts involving sources that are regarded as highly credible.

Belief systems and norms within groups that promote, enable and celebrate misogyny may be part of online-offline feedback loops that make sexual abuse appear more permissible – in both its online and offline forms – for those who are psychologically and situationally susceptible to influence of these kinds.

Under certain circumstances – such as times of significant uncertainty and "noisy" information ecosystems – individuals who are more likely to embrace ideological thinking may represent opportunitistic targets for gendered disinformation – as recruits to/supporters of an agenda and as spreaders of disinformation. Noisy information environments may stem from the multiple channels that characterize today's information ecosystem, or from deliberate tactics to "flood the [information] zone" with a constant barrage of provocations.

Because the misdirected beliefs, lack of empathy/compassion and the blame attached to the targets of online abuse involve elements of social learning, education and awareness can be seen as important elements of an effective response at the societal level as well as in relation to

opportunities for accountability, redress and rehabilitation by perpetrators (e.g., McIntyre, 2023; Cuklanz, 2023).

Public awareness, including campaigns by civil society actors to engage younger audiences and prospective male allies can play an important role in mainstreaming attention to, and action against, TF-GBV.

Social Media Literacy

The National Democratic Institute has proposed providing women with training to reduce online threats. This includes protecting personal information, using social media tools to minimize harassment, and strategies for maintaining mental health and resilience against online abuse (Jankowicz, et al., 2021). Literacy about the hazards of GD is discussed below in the section on mental immunity.

Ressa (2022) provided close-quarter insights into the design features of social media platforms (as described above). Knowledge of the ways that platform characteristic like the ease of re-posting content can contribute to its spread can help users be more prepared to resist these 'virality-by-design' features.

Platform literacy also involves understanding the broader factors at play. Ressa's (2022) experience sheds light on how platform business models can intersect with political and ideological agendas to target and harm women seen as threats to populist movements. This helps us better appreciate the challenges to constructive change. Addressing these issues requires active efforts from journalists, activists, civil society groups, government, and businesses that benefit from societal stability.

Implications for Countermeasures

Knowledge and training on the safe, informed, use of online platforms is advised to lower the risk of inadvertently falling prey to GD. This should be accompanied active messaging that online GD is not simply a personal problem or the result of the actions of those who have been impacted.

At the same time, harms flowing from some of the design features and macro-level dynamics of social media platforms and their business incentives cannot be mitigated by individual behaviour, alone. Addressing the broader risk environment will require multi-level, multi-modal action, with roles for government, civil society and businesses, as well as an informed citizenry.

Debunking: Exposure to Truths and the Viewpoints of Others

A fundamental threat – and objective – of gendered discrimination, and disinformation more generally, is the incitement of contempt for those who are defined as 'other'.

Nearly a century of social psychological research supports the idea that appropriate interpersonal contact[32] between diverse groups can improve relations by reducing perceived differences. However, this is effective only under certain conditions – that groups: share similar status and backgrounds; work cooperatively towards a common goal; and interact in a context that promotes positive cooperation and discourages division[33].

Online gendered disinformation lacks the situational features necessary for positive group interaction. However, research on the contact hypothesis helps us understand the role of situational determinants in behaviour. It suggests that factual corrections can counteract false beliefs arising from disinformation. This process, known as debunking, involves exposing and correcting false or misleading information to clarify the truth.

Debunking can be useful, but its effectiveness is limited. Research shows that misinformation tends to persist even after is corrected or withdrawn. This "continued influence effect" may occur because people tend to avoid the emotional cost of changing previously held beliefs (Susmann & Wegener, 2022).

Lewandowsky, et al. (2020) determined that effective refutations must clearly explain why information is false, and present the truth. Simply refuting a false fact is insufficient. Providing a credible alternative explanation or questioning the source's credibility can also be effective (Lewandowsky & van der Linden, 2021).

The source of the information is important. Debunking messages should come from individuals or organizations perceived as trustworthy by the audience (Lewandowsky, et al., 2020). However, if recipients ignore the source, its characteristics will have negligible effect.

If disinformation leads to misunderstandings about an issue, person or group, one might think that corrective narrative could help. However, this approach is often overly optimistic, as exposing people to contradictory information may actually reinforce their original beliefs.(Ecker, et al., 2020).

In an important study on debunking politically polarizing narratives, Bail, et al. (2018) conducted a field experiment with US Democrats and Republicans on Twitter. Participants were surveyed on policy positions and then exposed to periodic political content from the opposite party via bots. Contrary to expectations, instead of moderating initial views, exposure led to more polarized positions, especially among Republicans. Democrats demonstrated a slight, non-significant increase in liberal leanings. Bail, et al. (2018) suggested that, in light of the strong evidence from

[32] Widely known as the contact hypothesis, these ideas were articulated by Allport in 1954 as well as by Sherif and Sherif in the same year.

[33] Samson (2023); https://en.wikipedia.org/wiki/Contact_hypothesis

previous studies that inter-group contact can foster compromise and mutual understanding, future efforts to reduce political polarization on social media will likely need to focus on identifying the types of content or the positioning of messages that are prone to backfire, and whether other approaches and sources of information might be more effective.

Boukes and Hameleers (2023) explored the effectiveness of satire-based fact-checks as an alternative to traditional methods. They found that regular fact-based content reduced the perceived accuracy and credibility of false information, avoiding a backfire effect. By contrast, satire-based fact checks were found to be effective regardless of prior agreement with the fact-checked information. The researchers suggested that this may mean that refutations based on satire may be less vulnerable to resistance[34] and confirmation biases – possibly because of the cognitive effort required to participate in the narrative invoked by the satirical message.

However, neither approach decreased polarization based on commitment to group membership (in this study, political attitudes) and the gap between in-group like and out-group dislike (affective polarization[35]). Moreover, the use of satire was found to make it more likely that polarization would increase, whereas this was observed for regular debunking messages only when people saw the content of the refutations as confirming their existing views.

Implications for Countermeasures

Truth-restoring narratives and exposure to the views of others, by themselves, may be ineffective in counteracting disinformation. This is especially true – as is the case with GD – where one of the parties to the interaction does not perceive a sufficient degree of similarity with the target(s) of their attacks, where there is pre-existing polarization and when the context of the interaction favours polarization and conflict over harmony and cooperation.

There is emerging research suggesting that attempting to address online polarization simply by providing factual corrections to disinformation may actually exacerbate the problem through the production of backfire effects.

To stand a chance of effectively 'unsticking' disinformation narratives, debunking narratives must clearly articulate the ways that the narrative is incorrect. The counter narratives should lay out the details of the truth about the matter and be delivered by credible, trusted sources. However, the

[34] The case of resistance to belief is an interesting one. Individuals who prioritize self-direction, a human value that encourages independent thinking and actions, generally exhibit a higher need for cognition, whereas those who prioritize conformity, a value focused on maintaining the status quo, typically demonstrate a lower need for cognition (Coelho, Hanel & Wolf, 2020, cited in Kakinohana & Pilati, 2023). Thus, a disposition toward more thinking about the content of a message and its accuracy might be impacted by the cognitive energetic costs of considering the details of a satirical refutation.

[35] Boukes & Hameleers (2023); Bail (2021)

continued influence effect may render this unsuccessful. This is also the case when the content of fact-check messages are perceived as confirming existing views.

Satire-based fact-checks may be more broadly effective in reducing the perceived accuracy of a message and the credibility of the source. However, they may increase, rather than decrease affective polarization, either because it is seen as a critique of one's ideological identity or because of the cognitive load it imposes on people who might otherwise be able to process a message as false.

Upstream measures stand to be more effective than a more downstream response like debunking.

Forewarning and Prebunking: Psychological Inoculation to Disinformation

A promising for of upstream intervention is known as psychological inoculation. This can involve one or both of two components: (1) forewarning, which involves an advance caution to prepare recipients for the threat and motivate resistance; and (2) prebunking, a pre-emptive refutation of the anticipated message.

The continued influence effect causes false information to persist in memory even after convincing correction, as referencing the initial disinformation can reinforce its frame (Lewandowsky, et al., 2020). van der Linden (2021) proposed that instead of post-event debunking, using active-listening and focusing on misinformation techniques can be more effective. This approach leverages people's interest in avoiding manipulation.

A half-century of research on inoculation theory highlights techniques for building resistance to unwanted influence through protective exposure (McGuire, 1970, cited in Lewandowsky & van der Linden, 2021). Lewandowsky and van der Linden (2021) explain that by pre-emptively exposing individuals to a weakened form of manipulation, a cognitive-motivational process, similar to creating "mental antibodies," is triggered, enhancing resistance to future persuasion attempts.

Resistance is thought to be based in a mental default disposition to safeguard existing beliefs in the face of contradictory information. More recent scholarship on inoculation theory – emphasizing the virality of social media content – has expanded attention from narrow-spectrum, issue-specific, arguments toward a broader perspective including general influence and manipulation, as well as both active and passive approaches (Lewandowsky & van der Linden, 2021).

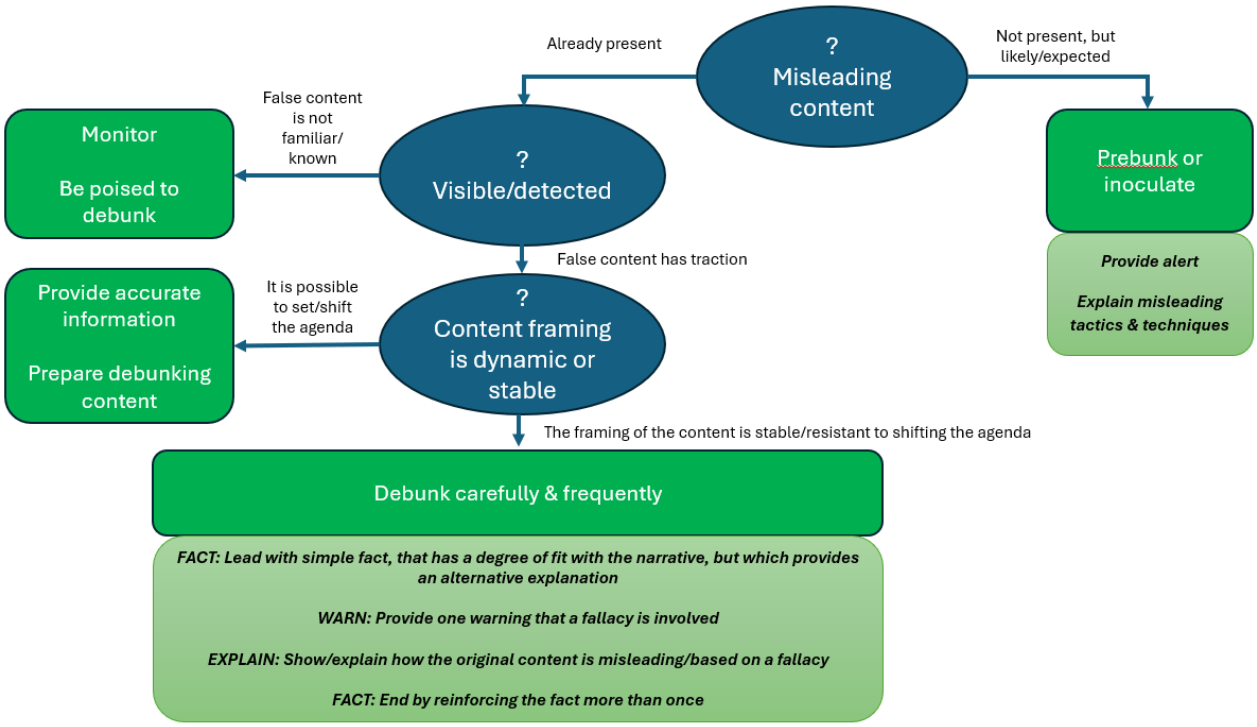
In contrast to the difficulty of counteracting conspiracy theories retrospectively through debunking, research shows that providing people with anti-conspiratorial content – either fact-based or logic-based – that foreshadows conspiracy theorist arguments can be effective (Lewandowsky & van der Linden, 2021). These techniques may also be effective against rhetoric used by online ideological extremists to radicalize prospective adherents to their cause.

A series of studies conducted by van der Linden and colleagues, summarized by van der Linden (2023) and Lewandowsky and van der Linden (2021) demonstrated that forewarning prospective target audiences about techniques of disinformation could neutralize the influence of false messages. This work also conformed to emerging evidence of an association between political ideology and susceptibility to untrue claims.

Research shows that individuals on the populist right may be more susceptible to disinformation and conspiracy theories than those on the ideological left (van der Linden, Panagopoulos, Azevedo, & Jost, 2020) – although both sides are susceptible under certain circumstances. These findings resemble those of Zmigrod's (2022) work on susceptibility to ideological thinking. They also reflect some of Bail's (2021) observations that ideologically conservative individuals may polarize further when exposed to opposing political content. Encouragingly, inoculation interventions can effectively moderate these impacts across target audiences.

Lewandowsky, et al. (2020) proposed a decision tree for determining when to deploy debunking or prebunking countermeasures (Figure 7). While this model focused on *mis*information, it has promise for misleading content more generally, including *dis*information. A more general application of this approach remains to be evaluated.

Figure 7. Decision tree for use of countermeasures (after model proposed by Lewandowsky, et al., 2020).



The key features of this decision tree recognize that prebunking has been found to be more effective than debunking. Consequently, debunking should be used only when necessary – that is, when false content has begun to gain traction and can no longer be ignored. In those cases, it must be done in structured way, and frequently, so that it has the best chance of competing with and displacing the false content through the power of repetition.

Using this formula, a **debunking** message might take the following form:

- **DISINFORMATION CLAIM THAT HAS STARTED TO GAIN TRACTION:** "Women aren't suited for leadership roles because they're too emotional to make rational decisions."
- **FACT:** Women are just as capable as men in leadership roles. Research shows that gender doesn't determine someone's ability to make sound decisions or lead effectively. In fact, many of the world's top-performing leaders are women.
- **WARNING:** Be careful—this kind of claim is based on a harmful stereotype, not facts.
- **EXPLAIN:** The idea that women are "too emotional" to lead is an outdated myth. It plays on old gender stereotypes and ignores real evidence. Emotional intelligence is actually a strength in leadership. Good leaders use both reason and empathy to make smart decisions. Misinformation like this can discourage women from taking on leadership roles and keeps unfair biases alive.
- **FACT (AGAIN):** There's no evidence that women are less effective leaders. Studies show that women perform equally well—or better—than men in leadership positions, across all sectors.

In the case of anticipated disinformation, based on the same false claim, a **prebunk** could include the following:

- **ANTICIPATED DISINFORMATION CLAIM TO GET AHEAD OF:** "Women aren't suited for leadership roles because they're too emotional to make rational decisions."
- **ALERT:** Heads-up! Be careful when you hear people say that women aren't good leaders because they're "too emotional." That's a stereotype designed to discredit women, not a fact.
- **EXPLANATION OF MISLEADING TACTICS:** What's really going on here – This kind of message uses an old stereotype to make women seem less capable. It wrongly suggests that showing emotion is a weakness, when actually, being in touch with emotions can help leaders connect, communicate, and make better decisions. These claims are meant to make people doubt women's abilities and stop them from taking on leadership roles.

To address the growing threat of online mis- and disinformation, Roozenbeek and van der Linden (2019) developed an innovative browser-based game called *Bad News*. The game, which is grounded in inoculation theory, was designed to build cognitive resistance to future disinformation attempts.

Roozenbeek and van der Linden (2019) observed significant improvements in the ability of participants to identify disinformation tactics after playing the game, with the strongest effects among those who were most susceptible to fake news before play started. This improvement was found across a range of demographic categories, including age, education level, political orientation, and gender.

Roozenbeek and van der Linden (2019) concluded that game-based inoculation can offer a "broad-spectrum psychological vaccine" without simply increasing overall skepticism. The latter point is important because one risk of awareness building around disinformation is that people will begin to see disinformation everywhere – leading to a generalized weakening of trust in the information environment and a greater likelihood of "false positives" – assuming that something is false when it is not.

Following the success of *Bad News*, the researchers developed several additional games, focusing on specific disinformation risks. Yet, despite the promise of using gamified approaches to inoculate people against disinformation, a degree of caution is warranted.

Using electronic games to build resistance to disinformation

Bad News is a short, interactive, simulation where players assume the role of a fake news producer. Over approximately 15 minutes, players work to gain followers and credibility by learning and applying six common misinformation tactics: impersonation; emotional manipulation; group polarization; conspiracy theory creation; discrediting sources; and trolling and baiting. Players earn badges for successfully applying these tactics in realistic social media-like scenarios. Ethical behavior is penalized in the game, reinforcing awareness of deceptive practices. The game was launched in partnership with a media outlet and was accessed by tens of thousands of users globally. A subset of over 14,000 participants completed pre- and post-game surveys to assess changes in their ability to recognize misinformation strategies.

Kiili, et al. (2024) conducted a systematic review of research, published between 2019 and 2021, focusing on the use of game-based and gamified learning environments designed to build skills for detecting false information. They identified that most of the 15 studies that matched their inclusion criteria reported positive outcomes for the interventions. However, they discovered considerable variation in what was measured and in the research designs used to assess effectiveness. They also observed that there is currently no standardized framework for describing and comparing across these types of techniques. As a

result, Kiili, et al. (2024) concluded that, not withstanding promising initial results, it is not yet possible to draw general conclusions about the effectiveness of these types of game-based interventions.

Maertens, et al. (2025) conducted a series of longitudinal experiments (with a total of 11,759 participants) to explore the persistence of misinformation resilience over time, following various inoculation interventions. They found that, while text-based and video-based interventions remained effective for up to one month, the effectiveness of game-based interventions – where the acquired skills may be cognitively more demanding to retain – decayed much more quickly. Importantly, they observed that booster interventions that are designed to enhance memory and recall of earlier learnings helped to offset the loss of effectiveness of all forms of interventions. The design of future research and interventions based on this work will likely focus on ways of enabling boosters to be designed and delivered at a pace conducive to sustaining the effective of counter-misinformation narratives. While this would likely be logistically complicated, advances in artificial intelligence – and potentially new features of social media platforms – could make this easier to achieve.

Implications for Countermeasures

Psychological inoculation is one of the most effective countermeasures against disinformation. It also avoids the risk of backfire effects known to be associated with debunking.

Moreover, forewarning about specific techniques of manipulation used in the context of certain topics, along with the refutation of anticipated messages (prebunking) appears to be effective in reducing differences in susceptibility to disinformation tied to divergent political ideologies.

Inoculation techniques may also help to lower the risk of radicalization to extremist ideological movements and so may represent an important upstream public health/public safety opportunity that benefits both prospective victims and prospective perpetrators.

Concepts of “mind parasites” and the mental immune system offer insights into the work that willful belief, distrust and cynicism do in signalling membership in identity-affirming groups. These ideas also underline the importance of broader societal efforts to provide persons vulnerable to radicalization with offramps towards more prosocial identities and their associated symbols and behaviours.

Inoculation measures are a natural fit with awareness and education campaigns and therefore, may be combined as a single package of upstream/prophylactic interventions. Gamified inoculations with booster interventions hold promise as a way to reach populations not always amenable to digital public health programs. However, more research will need to be done, including evaluation, in order to strengthen the empirical base for these efforts.

Contending with these phenomena will be challenging. The creation and propagation of mental immunity will require actions on many fronts to invite new experiences of belonging, and to supplant the symbolic and signalling systems of extremist groups with those of more moderate or prosocial coalitions.

Policy and Regulation: Potential Areas of Focus

While not, strictly speaking, countermeasures, attention to policy and regulatory options may be crucial to re-shaping the broader socio-cultural context towards more inclusive and less polarizing outcomes. Public concern about the prevalence of toxic content on social media has led to growing pressure on platforms to ensure more effective and accountable moderation (Sobieraj, 2020; Richardson-Self, 2021).

In the spyware industry, efforts to hold parties accountable, and to investigate or regulate them, can be challenging, due to complex and shifting ownership structures and corporate relationships (Marczak, et al., 2021). Citizen Lab researchers observed that many of these techniques are similar to those used by arms traffickers and money launderers (Marczak, et al., 2021).

Richardson-Self (2021) suggests enforcing clear standards and guidelines, to define speech identified as hate or abuse, and to require digital platforms to allocate resources to identify harmful content. However, the owners of major platforms such as Telegram and X are alleged to have resisted efforts to increase oversight (Mozur, et al., 2024). Richardson-Self (2021) also suggests user fees to slow the spread of harmful information, but notes this might simply drive users to other sites.

Ermoshina and Musiani (2025) propose implementing measures to ensure safer online spaces by design, analogous to Cavoukian's (2010) concept of privacy by design. Privacy by design encourages the view that privacy ought to be a core component of fair (and, ultimately, more effective) information practices – and essential to the functioning of democratic societies. Embodying seven foundational principles, privacy by design covers three main sets of applications: IT systems; accountable business practices; and physical design (Cavoukian, 2020).

Ermoshina and Musiani's (2025) "federated" model of content moderation offers an alternative to the top-down approaches used by major social media platforms. Instead of a single company setting the rules, this model is built on a network of independently run communities—each with its own moderation policies and user guidelines. Platforms, like Mastodon and Matrix, that use this approach, seek to allow communities to tailor their rules to local values and needs. Users can choose or move between communities that reflect their preferences, giving them more control over their online experience.

Ermoshina and Musiani (2025) suggest that such a decentralized model would support safer online spaces by encouraging moderation that is responsive, community-driven, and transparent. They

argue that it would also reduce the risk of one-size-fits-all policies and give people more say in how harmful content is handled.

While the federated model introduces models for user-centred, ethical, and flexible content moderation, it would also require ongoing investment in technical infrastructure, community participation, and shared responsibility to achieve its vision of supporting safer spaces. However, Ermoshina & Musiani (2025) suggest that, by promoting user choice and rejecting the profit-driven motives of large platforms, the federated model represents a promising path toward more ethical, inclusive, and accountable online environments.

Matthews (2021) assessed four main approaches to online content moderation drawing from private, community and legal models.

- **Legislation or government-led content moderation:** This top-down model involves the government defining what content must be moderated and by whom. It can take various forms, such as holding users accountable, tasking social media platforms with content removal, or establishing independent regulators. Legislation offers clarity and enforceability but may lack the flexibility to keep pace with technological change. Germany's Network Enforcement Act exemplifies this approach, placing the onus on platforms to remove illegal content within 24 hours, though it has faced criticism for encouraging over-censorship and giving too much power to private companies. Canada's proposed Online Harms Act (Bill C-63) draws on similar principles, though it is at a standstill.
- **Social network-led content moderation:** In this approach, social media companies take primary responsibility for moderating content, often driven by legal mandates or public pressure. While companies can tailor moderation to fit their platforms, critics argue that private sector control over speech poses risks to democratic discourse and transparency. There are also concerns about limited investment in moderation, ethics-washing, and monopolistic control due to the dominance of a few major platforms. Without transparency or oversight, users lack recourse when moderation decisions are made.
- **Third-party moderation tools:** This model promotes decentralization by enabling independent developers to build moderation tools that integrate with social media platforms—much like app stores in the tech industry. These tools offer users more choice and control over their online experiences and encourage competition. However, they face challenges such as unclear business models, privacy risks, and potential resistance from platforms. Nonetheless, Matthews (2021) concluded that tools like Block Party (for Twitter/X) demonstrate the potential for user-driven content control, particularly for communities most affected by online harassment.
- **Community-led moderation:** Community-led moderation is a bottom-up, pluralistic, competition-based approach where users set and enforce their own rules, often supported by platform tools and automation. Reddit is a leading example, empowering subreddit communities to self-govern within broad content guidelines. This model increases user

agency and diversity of experience but relies heavily on volunteer labor, raising concerns about sustainability, consistency, and the capacity for effective enforcement. Public responsibility on platforms like Reddit have not proven effective in preventing the development of highly misogynist online communities.

Lalonde, et al. (2025) argue that to improve transparency and accountability, holistic[36] and consistent platform policies should align with a practical regulatory regime than on corporate priorities. However, they are less optimistic about content moderation, as social media business models often hinder effective responses[37] to inappropriate content. Lalonde, et al.'s (2025) analysis of legal and policy responses to VMD – which aligns to the problem of gendered disinformation – is equally concerning. They conclude that, with growing technological sophistication, governments and international bodies are straining with how to regulate it effectively while respecting rights and adapting to evolving technologies.

For example, they point to international efforts by UNESCO and the UN to established non-binding principles to promote ethical AI use and digital platform governance which, however admirable, lack enforcement power.

Legislating against disinformation: A delicate balance

Addressing disinformation through legislation requires a careful balance – ensuring harmful content is identified and limited, while safeguarding forms of expression, such as satire, that play a legitimate and sometimes important role in exposing and challenging false narratives. In some cases, legislation may be used to suppress efforts to expose political agendas and activities which may, themselves, include inaccurate or misleading information.

To illustrate this challenge*, the Texas legislature recently passed a bill that would make it a crime to share altered political media – such as memes, videos, or audio – a unless it includes a government-approved disclaimer. Though originally intended to target AI-generated deepfakes, the legislation (House Bill 366) was expanded to cover any manipulated content that "did not occur in reality," including simple edits and parody. Despite recent amendments, the bill has drawn strong criticism in the US from First Amendment advocates, who argue it is overly broad and vague, potentially chilling political speech and satire. Questions remain about how the law would be applied.

Source: Waltens, B. (2025). Texas house approves former speaker Dade Phelan's meme regulation bill. *Texas Scorecard*, April 30, 2025. https://texasscorecard.com/state/texas-house-approves-former-speaker-dade-phelans-meme-regulation-bill/.

They assess that the European Union has taken the most proactive stance: the *Digital Services Act (DSA)* mandates risk assessments and algorithmic transparency by major platforms, while the *AI*

[36] For example, recognizing the presence and harms of both high-tech deepfakes and lower-tech "cheapfakes", and addressing policies to include a broader level of technological sophistication.

[37] Lalonde, et al. (2025) identify: removal – simple deletion of content; downranking – reducing content visibility by deprioritizing its position in search results and feeds; and demonetization – delinking online content from revenue generation.

Act requires clear labeling of synthetic content. In contrast, U.S. regulation remains fragmented, with states like California enacting laws targeting political deepfakes, but no unified federal framework exists. Domestically, Canada introduced the *Online Harms Act* (Bill C-63) to address AI-generated harms, especially to minors, but the bill stalled in early 2025, leaving a legal gap.

Lalonde, et al. (2025) conclude that, despite progress, there are significant challenges: legal fragmentation across jurisdictions weakens enforcement; the pace of technological change outpaces public policy and regulation; and the need to balance regulatory action with freedom of expression continues to pose dilemmas in democratic societies.

A 2018 report by the Public Policy Forum considered how to contend with the threats to democracy of harmful speech online (Tenove, et al., 2018). A key concern was that current regulations cannot tackle the massive and fast spread of harmful content across social media. One explanation offered is that foreign-owned platforms severely limit Canadians' ability to influence or oversee platform accountability. This creates a pronounced imbalance between the risks and the means available to Canadians to address them (Tenove, et al., 2018).

To help address this problem, the white paper outlined three interconnected public policy recommendations tailored to the Canadian context (Tenove, et al., 2018):

- **Adopt a multi-track framework for harmful speech regulation**
  A coordinated, multi-agency approach is needed to clarify how current laws can better address harmful online speech. This includes establishing a multi-agency task force, requiring social media companies to share data on harmful content with the public and researchers, and launching a multi-stakeholder commission to explore broader social and political impacts—fostering public dialogue on the future of content moderation and oversight.
- **Establish a moderation standards council**
  A new independent council—modeled after the Canadian Broadcast Standards Council—should be created to bring together platforms, civil society, and regulators. The Council would support transparent content moderation, develop and enforce codes of conduct, manage public complaints, and address regulatory conflicts, while contributing to international standards for online content governance.
- **Strengthen civil society and research capacity**
  Canada should significantly invest in research, programs, and civil society initiatives focused on harmful speech. Governments, academic institutions, and tech companies should support this work.

These recommendations are designed to work in a mutually supportive way to foster a healthier, more inclusive, and democratic digital public sphere in Canada. To the extent that they would be able to achieve this vision, each one would need to realize its full potential.

A more community-based approach involves the idea of coordinated acts of 'counterspeech' – speaking back against actions and systems that oppress people (Richardson-Self, 2021). The essence of this approach is to encourage collective action against harmful conditions and behaviours and greater accuracy by confronting biases and false assertions directly. This could take the form of refuting an inaccurate message and/or questioning the credibility of the source. In these ways, counterspeech is similar to notions of debunking discussed previously. Aligned to concerns identified earlier, Richardson-Self cites research by Costello and Hawdon (2020) that suggests that confronting hateful actors may only serve to amplify hateful rhetorics and their narratives.

Surveying the online regulatory landscape, Jankowicz, et al. (2024) assessed that not enough has been done by governments to mitigate online harms – either through incentives and requirements related to oversight, transparency and moderation, or through legislated responses such as criminalizing deepfake image based sexual abuse. Jankowicz, et al. (2024) offered eight recommendations addressing platform accountability and action and to address deepfake image-based sexual abuse. These recommendations covered the following areas:

- Government oversight of platforms to encourage improve duty of care related to women's ability to express themselves safely online;
- Transparency and oversight mechanisms enabling access by journalists and researchers to social media data, in the service of public interest, with appropriate privacy safeguards;
- Explicit provisions within online safety legislation and regulations to address online harms against women;
- Encouraging technology companies to address gender imbalances within their workforces;
- Institution of civil and criminal penalties for the creation and distribution of non-consensual deepfake pornography;
- Measures to interdict the availability and facility by which search engines websites and applications focused on the creation and distributions of deepfake pornography are used to harm women and children;
- Widening the availability of technologies that can be used to challenge deepfakes and protect original images from being misued ("immunizing images", digital "watermarks"); and
- Supporting public awareness campaigns and educational resources aimed at challenging deepfakes and remediating harms that have occurred.

An additional challenge concerns the difficulty of tracing an image back to the original upload. Robust and reliable technologies supporting correct attributions would be useful contibutors to both deterrence and accountability.

Implications for Countermeasures

Improved standards and guidelines are one element of a spectrum of higher-level responses to technology-facilitated harms against women and girls. However, these must be clear and practical, they must be properly resourced and implemented, and they must be transparent and enforceable through effective, accurate, reporting methods.

Legislation and regulations have been proposed to address the harms that flow from irresponsible or inadequately governed platforms. These may include civil and criminal penalties for non-compliance and for harms that stem from a lack of reasonable action by platform owners.

Calls for public funding for awareness and education about social media and gendered violence, including the use of deepfakes as methods of sexual abuse and exploitation, are consistent with the value of fostering awareness identified earlier.

Content moderation is perhaps the most broadly familiar measure in the public mind. Although there is likely a place for improved moderation, foreign ownership of social media platforms, deeper platform design features enabling virality, and business imperatives may lessen the impact of these types of interventions outside of coordinated, global, action by transnational coalitions.

Developing a capacity for reliable and valid attribution would play an important role in deterrence.

Support for Those Affected by Gendered Disinformation

Women involved in politics – especially women of colour – face repeated and ongoing online violence (Sobieraj, 2020). The National Democratic Institute has urged that social media platforms should have specific contacts to whom reports of online abuse could be escalated (Jankowicz, et al., 2021).

Jankowicz, et al. (2024) recommend that, in addition to investing in public awareness campaigns and the development of educational resources about technology-facilitated gender-based violence, governments should also support capacity building within the police and justice communities to enable them to be able to respond more effectively to enforcement situations and community building opportunities. Similarly, they recommend that schools and employers have policies and supports in place for students and staff who experience TF-GBV.

Sobieraj (2020) emphasizes that this is a fundamentally anti-social and anti-democratic phenomenon. Consequently, collective action and a network of support will be necessary to help individuals recover from harms that have occurred and to experience opportunities for resilience in the midst of ongoing threats. More importantly, to create lasting change that inoculates society against the threats of online misogyny, macro-level attention, joined-up action, and more active roles for governments, community-based organizations and digital platforms – in keeping with concerns related to privacy, free-speech and due process – will be necessary.

To bring the focus further upstream, it will also be necessary to address structural, socio-economic and other macro-features of our communities and society that create conditions of risk. These include developmental traumas and experiences impacting boys and young men, which constitute the conditions of exclusion and despair that are grist for the narrative mill of misogynistic authoritarian movements, as suggested earlier.

Implications for Countermeasures

Services and supports for those who have been victimized, or are recovering from, technology-facilitated gender based violence are an important component of a holistic and shared response to this problem.

This can be enhanced by training for the justice, educational and community sectors focusing on strengthening their individual and collective capacities to prevent and intervene in the aftermath of online abuse.

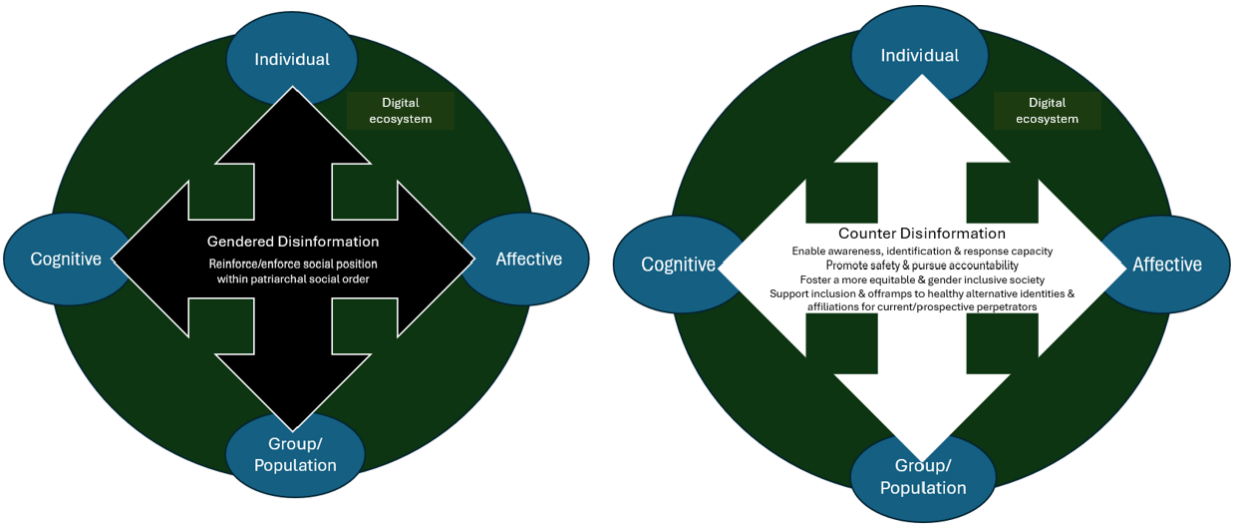
In addition to downstream supports, mid-stream and upstream measure focusing on enhancing conditions that community safety and wellbeing, will serve better developmental outcomes for children, and make communities lower in social determinants of risk and richer in social determinants of wellbeing.

A Strategy for Change

Mindful of the foregoing research, concerns and caveats about the scope and complexity of gendered disinformation, it is important to start with an eye to building momentum and fostering networked capacity for contending with this problem. A strategic mix of countermeasures that span the upstream-midstream-downstream continuum would help shape conditions that are: more resistant to misogyny and disinformation, less conducive and more responsive to technology-facilitated violence against women (whether perpetrated as individual acts of misogyny or as tools of foreign interference), and more supportive of the resilience and recovery of those targeted by GD.

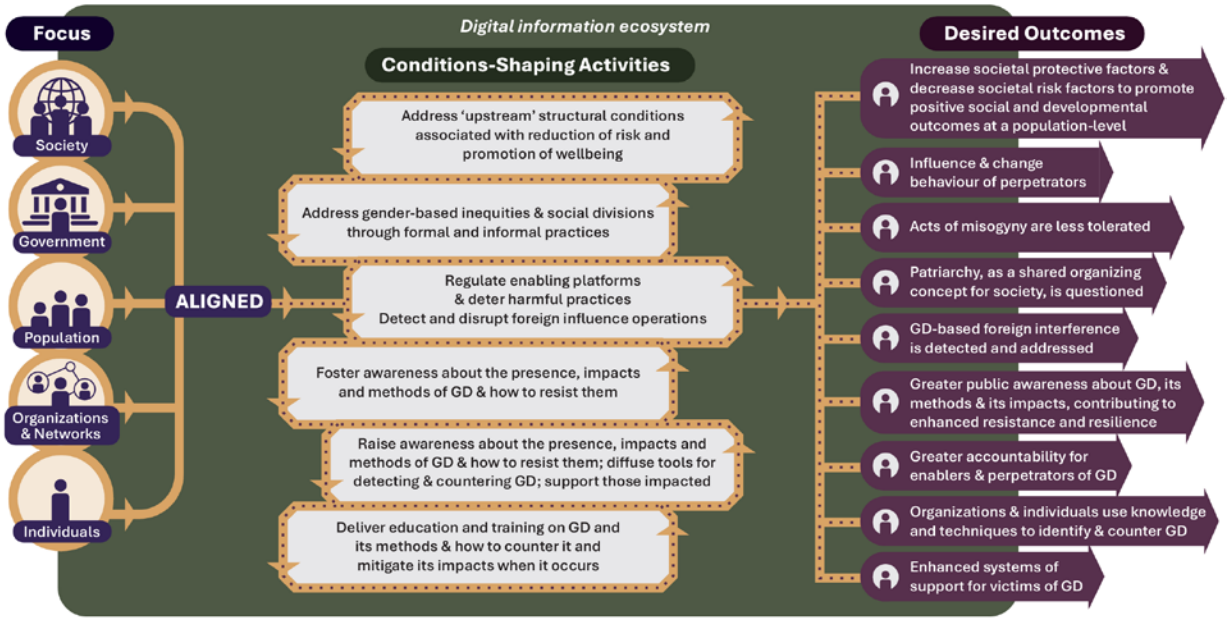
A suite of effective countermeasures should start with attention to awareness, providing skills, tools and opportunities for support to those who have been affected, and work along the length of the intervention continuum (Figure 8).

Figure 8. Focus of gendered disinformation and counter disinformation activities.



Gendered disinformation occurs within a broad and varied socio-cultural context. A corresponding theory of change for addressing GD as a holistic, all of society problem, is shown below (Figure 9). It involves efforts at multiple levels that, if aligned, would create a set of mutually reinforcing conditions that increase the probability of realizing a constellation of desired outcomes.

Figure 9. Preliminary theory of change for addressing gendered disinformation holistically as an all-of-society problem.

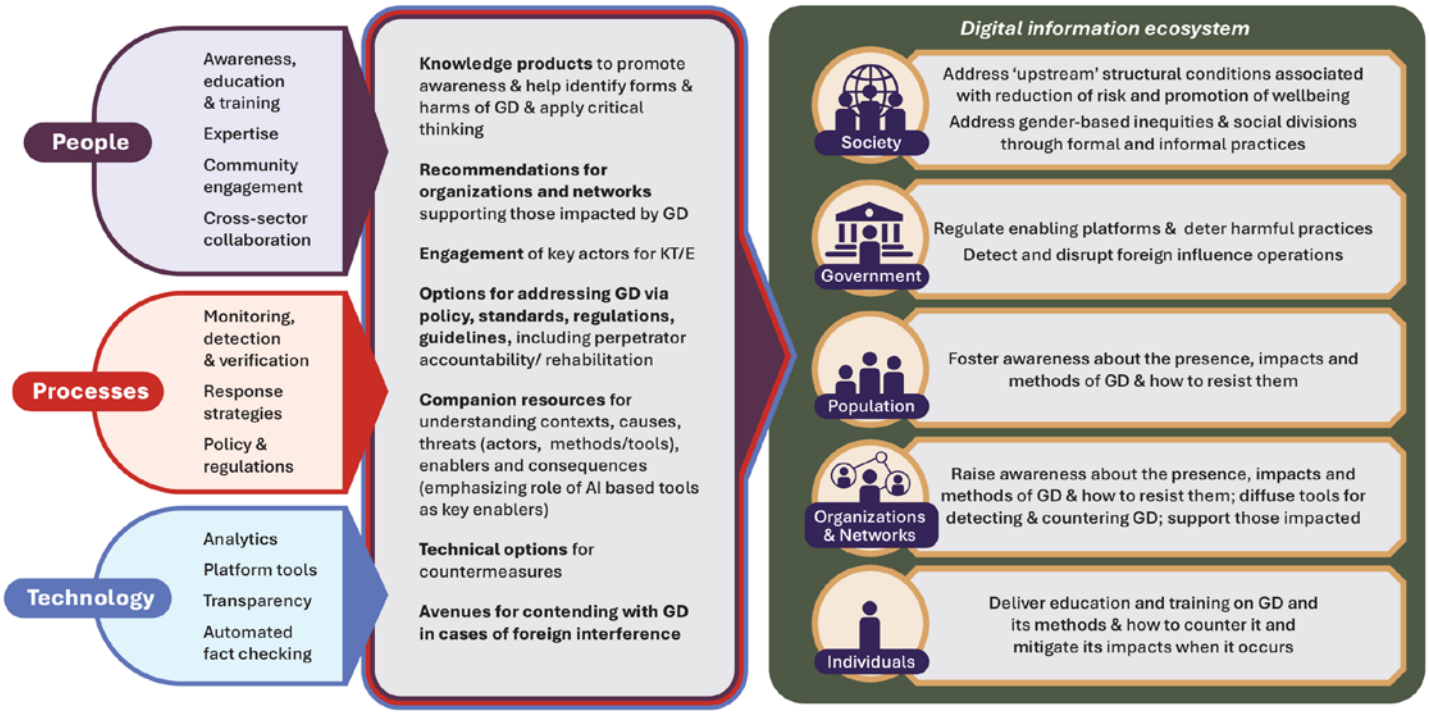


At this juncture, the most promising avenues appear to be those that address: awareness and identification; response capacity; support and empowerment; and policy engagement. The components of this multi-level approach emerge as a basis for achieving initial traction against what has seemed to be an intractable problem. On the basis of the preceding discussion, several considerations stand out:

- Awareness-based interventions are critical for building resistance to GD and resilience enhancing networks of support and accountability.
- Tapping into knowledge-building and training opportunities within the educational and human service systems may help to reduce participation in, and victimization by, malicious information exploits among young people and foster improved response capacity among educators, police, public health and community partners. Awareness building and training involving police and community partners might build on work already underway on the related topics of intimate partner violence and coercive control (e.g., Gill, et al., 2021). Awareness and educational opportunities for school-aged populations might be incorporated into a range of curricular learnings touching on information technology, AI and cybersafety (e.g., social sciences, humanities, computer science) and extracurricular activities focusing on women in science, technology, engineering and math (STEM).
- A focus on gender as a tool of foreign interference and authoritarianism may help the policy, national security and law enforcement communities in their efforts to identify and counter FIMI exploits as well as domestic ideological movements engaging in stochastic terror practices.
- Highlighting how social media platforms help spread GD can increase public understanding and support for policies that balance risk mitigation with freedom of expression.
- Wherever feasible and appropriate, interventions should be tailored to local contexts, considering factors such as literacy, access to technology, and existing gender norms.
- It is not enough to focus on individuals alone as this is not only an individual trouble; it is a collective threat.
- Interventions at the individual level should be combined with organizational, network, technological and policy solutions for maximum effectiveness, as part of a comprehensive approach to combatting gendered disinformation.

A corresponding system for countering gendered disinformation, consisting of people, processes and technology, is summarized below (Figure 10).

Figure 10. Counter GD system of people, processes and technology.



Factors considered in developing this system – which should be considered in its implementation, include the following.

- **Practical utility:** The system should address the widest array of GD threat scenarios.
- **Perceived relevance and value:** The system should address perceived needs across a broad spectrum of users (organizations/agencies, government, individuals) and should serve as a basis for engaging prospective users on needs that are real but, not yet, experienced.
- **Adaptability and maintenance:** The system should allow for flexibility in use and should be able to be updated as new information becomes available (e.g. threat sharing) and/or as threats and the enabling technologies continue to evolve.
- **Capability and cost:** The system should be accessible to a wide range of users, with advanced users able to gain more benefits than those with less technical expertise.

A detailed overview of this system is provided in Appendix C. Appendix D provides a curated set of sample technologies that might be useful to individuals, and human service and educational organizations in identifying and countering potential instances of gendered disinformation.

Appendix E includes a list of the set of accompanying knowledge resources to support awareness and actions among: educators (Annex E1); families and youth (Annex E2a, E2b); a set of additional resources for educators, families and youth (Annex E2c); police and community partner agencies (Annex E3); and government stakeholders (Annex E4). These resources are contained in the companion document, *Understanding and Countering Gendered Disinformation: Knowledge Resources*.

Included within Annex E4 are a set of recommendation to support the development of an expanded national capacity for contending with gendered disinformation. These recommendations follow.

CONCLUSION

Gendered information is a complex issue linked to polarization, patriarchy and misogyny – driven by individuals, groups and nation states. It targets women, girls and gender non-conforming persons, causing harm as victims or tools of repression. No single approach can counter these threats to safety and national security. Therefore, countering gendered disinformation requires a multi-layered, strategic framework that promotes awareness and builds a networked response capacity. It should focus on strengthening resistance to disinformation, and gendered information specifically. Because of the shared nature of these threats, this should involve joined-up coordinated efforts to prevent and mitigate risk, foster resilience, and balance solutions with our democratic values.

Gendered disinformation about Indigenous women and girls is deeply rooted in Canada's colonial history and current realities, with serious consequences. Corbett (2019) observed that inaccurate portrayals in media and culture reinforce negative stereotypes, leading non-Indigenous Canadians to ignore ongoing violence. Corbett recommends breaking this cycle by challenging false narratives, changing harmful media practices, and prioritizing Indigenous voices in storytelling. Together, these measures can help change the harmful information landscape and support reconciliation.

The proposed system emphasizes multi-sectoral collaboration involving people, processes and technology. Knowledge development will be essential to building networked capacity to counter gendered disinformation. This should offer mutual benefits and support shared learning, planning and implementation.

We propose a theory of change involving strategically aligned, society-wide interventions grounded in emerging research. We also offer a set of knowledge resources and technology examples useful to those in human services, policy and national security. Finally, we recommend creating a cross-sectoral knowledge development and mobilization network to support evidence-informed, collaborative efforts on this important issue.

RECOMMENDATIONS

Policy, Legislation and Enforcement

2. That the federal government:

    a. Implement policy and legislative measures to counter gendered disinformation, recognizing that it is a threat that spans community safety and wellbeing, and national security.
        - *The corresponding regulatory framework should ensure platform accountability, transparency, and meaningful financial penalties for non-compliance.*

    c. With targeted investment, initiate cross-departmental, industry, academic and private sector operational coordination and program collaboration to address gendered disinformation within public safety, public health, digital regulation, defence and national security frameworks.

    d. Develop a national strategy on gendered disinformation in close partnership with the private sector, research and civil society, integrating public safety, digital governance, and foreign policy approaches.

    j. Convene and engage women's advocacy organizations, racial justice groups, security and intelligence professionals, academic researchers, cyber-security experts and relevant community and private sector entities in dialogue on such matters as how to optimize the balance of protection and enforcement with freedom of expression online.

    k. Increase data collection and monitoring of gendered disinformation trends and actionable current intelligence.

    l. Conduct periodic cross-sector consultations with experts in gender-based violence, cybersecurity, open source intelligence, national security, and digital regulation to understand the evolving landscape of gendered disinformation.

m. Establish gender-responsive online safety laws that hold technology platforms accountable. Options include the re-introduction of Bill C-36[38] and the applications of relevant elements of a Clean Pipes Strategy[39].

n. Enhance training for security, intelligence, diplomatic, defence, law-enforcement and policymakers on technology-enabled GD.

o. Invest in digital literacy, research, open source intelligence and enforcement mechanisms to strengthen Canada's resilience against gendered disinformation.

Research and Knowledge Mobilization

3. That Canada support the creation of a cross-sectoral knowledge mobilization network on gendered disinformation – the Gendered Disinformation Knowledge Network (GenD-Net).

   Such a network would serve as a hub for leadership, information sharing, education and training, research, and policy coordination, program planning, operational coordination and de-confliction ensuring that responses to gendered disinformation are evidence-based, and aligned across sectors.

   The objectives of the network will be to:

[38] Canada's Bill C-36 (proposed) sought to amend hate speech provisions to better address online harms, including gender-based hate. Canada's Online Harms Act (Bill C-63), officially titled, *An Act to enact the Online Harms Act, to amend the Criminal Code, the Canadian Human Rights Act and An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service and to make consequential and related amendments to other Acts*, aimed to address harmful content on the internet. In particular, issues related to child exploitation, hate speech, and content promoting violence or self-harm. The Bill would establish a Digital Safety Commission to oversee compliance, investigate complaints, and enforce penalties. The Bill also aims to hold platforms accountable for the content that is hosted on their platform. In particular, it creates several Duties on the platform such as a duty to act responsibly, protect children and keep all the records. If the Bill were to receive Royal Assent, then the legislation would increase penalties for hate crime, expand the definition of hate crime and amend elements of the Criminal Code of Canada. It needs to be noted that the Bill was not passed prior to the 2025 election, hence, it is currently not codified in law.

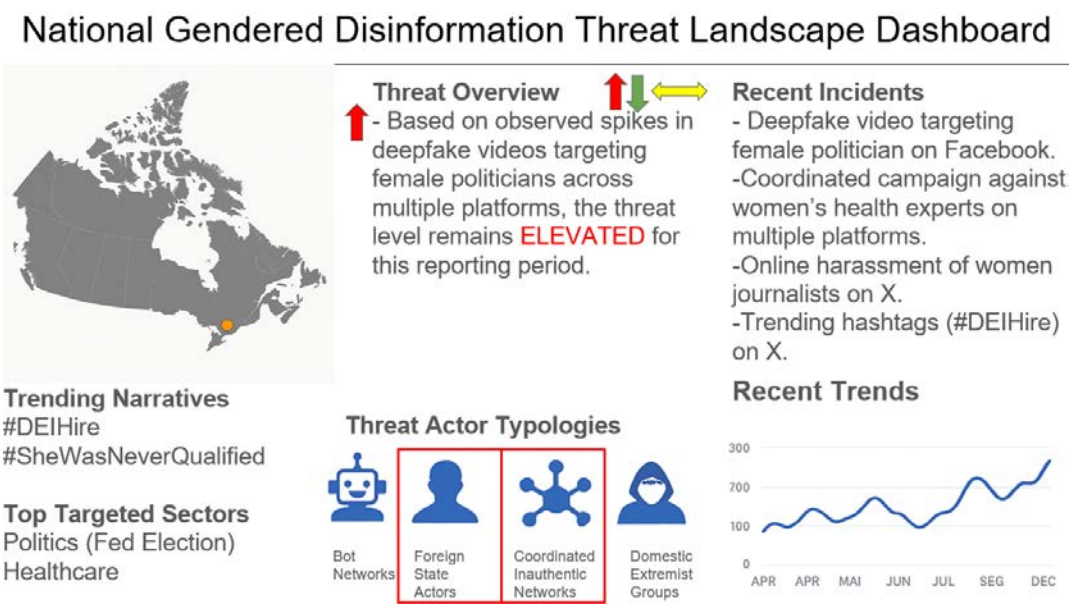
[39] A "clean pipes" strategy is a cybersecurity approach where internet service providers filter out malicious traffic—such as malware, phishing, and botnet activity—before it reaches end users. By blocking known threats at the network level, it helps create a safer online environment and reduces the burden on individuals and organizations to defend themselves. This strategy is part of national cybersecurity efforts in several countries, including the United Kingdom, Australia, and Singapore, which have partnered with internet service providers (ISPs) to implement network-level threat filtering to protect citizens and critical infrastructure.

- *Enhance knowledge mobilization and public awareness of gendered disinformation.*
- *Support curriculum development, stimulate and contribute to education and training.*
- *Strengthen community and cross-sectoral dialogue and collaboration on policy development.*
- *Support defence, intelligence, police and public safety agencies.*
- *Advance research and innovation, including evaluation capacity building.*
- *Bridge gaps in service provision for affected communities.*

Gendered Disinformation as a National Security Issue

4. That the Government of Canada refine and implement options for countering gendered disinformation as a national security issue, including its use as an element of foreign interference. Enhance the capabilities of defensive cyber operations in relation to this threat. More particularly:
    g. Establish a dedicated government funding stream for research and innovation on gendered disinformation that is open to Canadian industry, academia and not-for profit organizations.
    h. Incentivize Canadian industry participation and innovation through public-private partnerships and direct investment.
    i. Develop a national strategy on gendered disinformation as a foreign interference threat, and ensure integration with national defence policy, cyber security and national security strategies.
    j. Fund the creation of a cross-sectoral intelligence-sharing network to combat gendered disinformation, including the creation and maintenance of a national gendered disinformation threat landscape reporting capacity; this would, in-turn, feed into an intelligence "dashboard" (Figure 11) which could be made publicly available as part of building overall awareness an public will to confront this problem (See Annex E4, Attachment B).
    k. Establish legal and policy frameworks to protect women in public life from both foreign and domestic online harm.
    l. Develop a rapid response mechanism to protect individuals facing high-risk disinformation attacks (see Annex E4, Briefing Resources 1 and 4).

Figure 11. Sample depiction of a proposed gendered disinformation dashboard.



Impact of Recommendations

Implementing these recommendations will have significant impacts on combatting gendered disinformation, enhancing human rights protection, and promoting gender equality. By addressing this issue, intertwined with polarization and misogyny, we can safeguard women, girls, and gender-nonconforming individuals from targeted harm. More specific areas impacted are as follows:

Policy and Legislation

By implementing comprehensive policies and legislation, the federal government will strengthen community safety and national security. Establishing regulatory frameworks with platform accountability and penalties for non-compliance will ensure that digital spaces are safer and more transparent. Cross-departmental coordination will enhance efforts to address gendered disinformation within public safety and national security frameworks.

Multi-Sector Collaboration

Creating a national strategy in partnership with the private sector, research institutions and civil society will integrate approaches to enhancing both public safety and social media governance. Engaging diverse organizations in dialogue will balance safety and security with freedom of expression. Furthermore, this approach will help build resilience against gendered disinformation through enhanced data collection, training, and digital literacy investments.

Research and Knowledge Mobilization

A dedicated funding stream for research and innovation, alongside public-private partnerships, will drive industry participation and technological advancements.

Establishing the Gendered Disinformation Knowledge Network (GenD-Net) will enhance public awareness, support curriculum development, and foster cross-sectoral collaboration. By bridging gaps in service provision, it will ensure evidence-based responses aligned across sectors.

National Security

Recognizing gendered disinformation as a national security issue will help refine strategies to counter foreign interference. Developing a rapid response mechanism and legal frameworks will protect individuals from high-risk disinformation attacks.

Overall, when implemented, these measures will help to transform the online information landscape, support reconciliation, and uphold Canadian liberal democratic values by fostering a coordinated, strategic response to gendered disinformation.

REFERENCES

Ai Ramiah, A. & Hewstone, M. (2013). Intergroup contact as a tool for reducing, resolving, and preventing intergroup conflict evidence, limitations, and potential. *American Psychologist*, 68(7):527-542. DOI: 10.1037/a0032603.

Aljizawi, N., Anstis, S., Michaelsen, M., Arroyo, V., Baran, S., Bikbulatova, M., Böcü, G., Franco, C., Geybulla, A., Iliqud, M., Lawford, N., LaFlèche, E., Lim, G., Meletti, L., Mirza, M., Panday, Z., Posno, C., Reichert, Z., Taye, B., & Yang, A. (2024). *No escape: The weaponization of gender for the purposes of digital transnational repression* (Citizen Lab Report No. 180). University of Toronto. https://citizenlab.ca/2024/12/the-weaponization-of-gender-for-the-purposes-of-digital-transnational-repression/.

Bail, C.A. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton, NJ: Princeton University Press.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Fallen Hunzaker, M.B., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Procedings of the National Academy of Science*, 115(37), 9216-9221. https://www.pnas.org/doi/epdf/10.1073/pnas.1804840115.

Barker-Singh, S. (2025). MP tells Sky News she was targeted online by Tate brothers after Commons contribution. *Sky News*, April 3, 2025. https://news.sky.com/story/mp-tells-sky-news-she-was-attacked-online-by-tate-brothers-after-commons-contribution-13340655.

Besancenot, M-D. (2025). Next time you hear someone say “it’s just coms”, pull out the striking visuals provided by the European External Action Service (EEAS) in their last report on information threats. LinkedIn post, March 26, 2025. https://www.linkedin.com/feed/update/urn:li:activity:7310592850847559680/.

Biddlestone, M., Azevedo, F. & van der Linden, S. (2022). Climate of conspiracy: A meta-analysis of the consequences of belief in conspiracy theories about climate change. *Current Opinion in Psychology*, 46, 101390. https://doi.org/10.1016/j.copsyc.2022.101390.

Bijlsma, A. M. E., van der Put, C. E., Vial, A., van Horn, J., Overbeek, G., & Assink, M. (2022). Gender differences between domestic violent men and women: Criminogenic risk factors and their association with treatment dropout. *Journal of Interpersonal Violence, 37*(23-24), NP21875-NP21901. https://doi.org/10.1177/08862605211072704

Boukes, M. & Hameleers, M. (2023) Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation. *Communication Monographs*, 90(1), 69-91, DOI:10.1080/03637751.2022.2097284.

Bradshaw, S. & Henle, A. (2021). The gender dimensions of foreign influence operations. *International Journal of Communication*, 15(2021), 4596-4618. https://ijoc.org/index.php/ijoc/article/view/16332/3584.

Canadian Women's Foundation (n.d.). *The facts about gendered digital hate, harassment, and violence*. https://canadianwomen.org/the-facts/online-hate-and-cyberviolence/.

Cavoukian, A. (2010). *Privacy by design: the definitive workshop*. A foreword by Ann Cavoukian, Ph.D. *IDIS, 3*, 247–251. https://doi.org/10.1007/s12394-010-0062-y.

CCA (Council of Canadian Academies). (2023). *Vulnerable Connections*. Ottawa (ON): Expert Panel on Public Safety in the Digital Age, CCA. https://cca-reports.ca/wp-content/uploads/2023/04/Vulnerable-Connections_FINAL_DIGITAL_EN_UPDATED.pdf.

Coelho, G.L.H., Hanel, P.H.P. & Wolf, L.J. (2020). The very efficient assessment of need for cognition: developing a six-item version. *Assessment*. 27(8):1870–1885. doi: 10.1177/1073191118793208.

Corbett, E. (2019). When disinformation turns deadly: The case of missing ad murdered Indigenous women and girls in Canadian media. In J. McQuade, T. Kwok & J. Cho (Eds.), *Disinformation and digital democracies in the 21st century*. Toronto, ON: The NATO Association of Canada, 19-23.

Costello, M. & Hawdon, J. (2020). Hate speech in online spaces. In T. Holt & A. Bossler (Eds.), *The Palgrave handbook of cybercrime and cyberdeviance*. London: Springer Nature, 1397-1416.

Dawson, M., Sutton, D., Carrigan, M., Grand'Maison, V., Bader, D., Zecha, A., & Boyd, C. (2019). *#CallItFemicide: Understanding gender-related killings of women and girls in Canada 2019*. Canadian Femicide Observatory for Justice and Accountability. https://femicideincanada.ca/callitfemicide2019/pdf.

Deibert, R. (2025). *Chasing shadows: Cyber espionage, subversion and the global fight for democracy*. New York: Simon & Schuster.

Dewey, C. (2014). Inside the 'manosphere' that inspired Santa Barbara shooter Elliot Rodger. *The Washington Post*, May 27, 2014. https://www.washingtonpost.com/news/the-intersect/wp/2014/05/27/inside-the-manosphere-that-inspired-santa-barbara-shooter-elliot-rodger/.

DiMeco, L. (2019). Gendered disinformation, fake news, and women in politics. *Council on Foreign Relations*, December 6, 2019. https://www.cfr.org/blog/gendered-disinformation-fake-news-and-women-politics.

Douglas, H., Harris, B. & Dragiewicz, M. (2019). Technology-facilitated domestic and family violence: Women's experiences. *British Journal of Criminology*, 59, 551-570. doi:10.1093/bjc/azy068.

Ecker, U. K. H., Lewandowsky, S., & Chadwick. M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5, 41. https://doi.org/10.1186/s41235-020-00241-6.

Economist Intelligence Unit (2020). *Measuring the prevalence of online violence against women*. https://onlineviolencewomen.eiu.com/.

Equal Measures 2030 (2024). *A gender equal future in crisis? Findings from the 2024 SDG Gender Index*. Seattle, WA: Equal Measures 2030.

Ermoshina, K. & Musiani, F. (2025). Safer spaces by design? Federated socio-technical architectures in content moderation. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1827.

Fazio, L.K, Brashier, N.M., Payne, B.K. & Marsh, E.J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993-1002.

Fazio, L.K. & Sherry, C.L. (2020). The effect of repetition on truth judgments across development. *Psychological Science*, 31(9), 1150-1160.

French, A.M., Storey, V.C. & Wallace, L. (2025). The impact of cognitive biases on the believability of fake news. *European Journal of Information Systems, 34*(1), 72-93, DOI: 10.1080/0960085X.2023.2272608.

Gill, C. & Aspinall, M. (2020). *Understanding coercive control in the context of intimate partner violence in Canada*. Research paper for the Office of the Federal Ombudsman for Victims of Crime, Department of Justice Canada. Fredericton, NB: University of New Brunswick.

Gill, C., Campbell, M.A. & Ballucci, D. (2021). Police officers' definitions and understandings of intimate partner violence in New Brunswick, Canada/ *The Police Journal, 94*(1), 20-39. https://doi.org/10.1177/0032258X19876974.

George, J., Gerhart, N., & Torres, R. (2021). Uncovering the truth about fake news: A research model grounded in multi-disciplinary literature. *Journal of Management Information Systems, 38*(4), 1067–1094. https://doi.org/10.1080/07421222.2021.1990608.

Hameleers, M. (2022). *Populist disinformation in fragmented information settings: Understanding the nature and persuasiveness of populist and post-factual communication*. London: Routledge.

Hasher, L., Goldstein, D. & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107-112.

Hess, A. (2014). Why women aren't welcome on the Internet. *Pacific Standard*, January 6, 2014. https://psmag.com/social-justice/women-arent-welcome-internet-72170/.

Human Rights Watch. (2024). *We will find you: A global look at how governments repress nationals abroad*.

https://www.hrw.org/sites/default/files/media_2024/02/global_transnationalrepression0224web_0.pdf

Hutchins, E., Cloppert, M.J. & Amin, R.M. (2010). *Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains*. Unpublished report. Lockheed Martin. https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf

Jankowicz, N., Pepera, S. & Middlehurst, M. (2021). *Addressing online misogyny and gendered disinformation: A how-to guide*. National Democratic Institute. https://www.ndi.org/sites/default/files/Addressing%20Gender%20%26%20Disinformation%202%20%281%29.pdf

Jankowicz, N., Gomez-O'Keefe, I., Hoffman, L. & Vidal Becker, A. (2024). *It's everyone's problem: Mainstreaming responses to technology-facilitated gender-based violence*. New York, NY: Columbia University SIPA Institute of Global Politics and the Vital Voices Global Partnership. https://igp.sipa.columbia.edu/sites/igp/files/2024-09/IGP_TFGBV_Its_Everyones_Problem_090524.

Kakinohana, R.K. & Pilati, R. (2023). Differences in decisions affected by cognitive biases: examining human values, need for cognition, and numeracy. *Psicologia Reflexao e Critica*., *36*(1):26. doi: 10.1186/s41155-023-00265-z.

Kelshall, C. (2020). Soft violence, social radicalisation and violent transnational social movements (VTSMs). Paper presented at the November 25, 2020 meeting of the CASIS West Coast Security Conference, Vancouver, BC. *Journal of Intelligence, Conflict and Warfare, 3*(3). https://doi.org/10.21810/jicw.v3i3.2800.

Kesivan, M. (2024). India is witnessing the slow-motion rise of fascism. *The Guardian*, September 8, 2024. https://www.theguardian.com/commentisfree/article/2024/sep/08/india-slow-motion-rise-of-fascism?CMP=Share_iOSApp_Other.

Kiili, K., Siuko, J. & Ninaus, M. (2024). Tackling misinformation with games: a systematic literature review. *Interactive Learning Environments, 32*(10), 7086-7101, DOI: 10.1080/10494820.2023.2299999.

Kolga, M. (2024, October 16). *Testimony before the Canadian House of Commons Standing Committee on Public Safety and National Security, October 1, 2024*. Macdonald-Laurier Institute. https://macdonaldlaurier.ca/marcus-kolga-warns-against-threat-of-russian-cognitive-warfare-mli-in-parliament/.

Korteling, J.E. & Toet, A. (2020). Cognitive biases. In S. Della Sala (Ed.), *Reference Module in Neuroscience and Biobehavioral Psychology*. Amsterdam-Edinburgh: Elsevier ScienceDirect. https://doi.org/10.1016/B978-0-12-809324-5.24105-9.

Lalonde, M., Boulianne, G., Rutherford, N., Beaulieu, M., Ghodrati, H. & Dahmane, M. (2025). *Visual and multi-modal disinformation: Analysis, challenges, solutions*. Montreal, QC: Computer Research Institute of Montreal (CRIM) & Ottawa, ON: Information Integrity Lab, University of Ottawa.

Lanier, J. (2018). *Ten arguments for deleting your social media accounts*. New York, NY Holt.

Lewandowsky, S., Cook, J., Ecker, U., Albarracín, D., Kendeou, P., Newman, E.J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G., Swire-Thompson, B., van der Linden, S., Wood, T.J., & Zaragoza, M. S. (2020). *The debunking handbook 2020*. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1247&context=scholcom.

Lewandowsky, S. & van der Linden, S. (2021) Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, DOI: 10.1080/10463283.2021.1876983.

Lilly, M. 2016. *The world is not a safe place for men: The representational politics of the manosphere*. Unpublished masters thesis, University of Ottawa. https://ruor.uottawa.ca/server/api/core/bitstreams/1eee5112-7f22-4ffc-a49d-a978a56bed05/content.

Maertens, R., Roozenbeek, J., Simons, J.S., Lewandowsky, S., Maturo, V.,, Goldberg, B., Xu, R. & van der Linden, S. (2025). Psychological booster shots targeting memory increase long-term resistance against misinformation. *Nature Communications, 16*, 2062 (2025). https://doi.org/10.1038/s41467-025-57205-x.

Maimann, K. (2024). Instagram ignored 93% of abusive comments toward female politicians: Report. *CBC Online News*, August 19, 2024. https://www.cbc.ca/news/women-politicians-online-abuse-1.7298168.

Marczak, B., Scott-Railton, J., Razzak, B.A., Al-Jizawi, N., Anstis, S., Berdan, K. & Deibert, R. (2021). *Pegasus vs. Predator: Dissident's doubly-infected iPhone reveals Cytox mercenary spyware*. The Citizen Lab Research Report No. 147, University of Toronto, December 2021. https://citizenlab.ca/2021/12/pegasus-vs-predator-dissidents-doubly-infected-iphone-reveals-cytrox-mercenary-spyware/

Maté, G. (2022). *The myth of normal: Trauma, illness and healing in a toxic culture*. Toronto, ON: Knopf Canada.

Maté, G. (2024). We each have a Nazi in us. We need to understand the psychological roots of authoritarianism. *The Guardian*, September 6, 2024. https://www.theguardian.com/commentisfree/article/2024/sep/06/authoritarianism-roots-origin?CMP=Share_iOSApp_Other.

Matthews, M. (2021). *Four approaches to content moderation and their risks and benefits*. Information and Communications Technology Council (ICTC), October 20, 2021. https://ictc-ctic.ca/articles/four-approaches-to-content-moderation-and-their-risks-and-benefits.

McIntyre, L. (2023). *Post-truth*. Cambridge, MA: MIT Press.

McIntyre, L. (2023). *On disinformation: How to fight for truth and protect democracy*. Cambridge, MA: MIT Press.

McMahon, D. (2021). *Cyber deception: The art of camouflage, stealth and misdirection*. Unpublished paper. Ottawa, ON: Clairvoyance Cyber Corp.

Michaelsen, M. & Anstis, S. (2025): Gender-based digital transnational repression and the authoritarian targeting of women in the diaspora, *Democratization*, DOI: 10.1080/13510347.2025.2476178.

Mozur, P., Satariano, A., Krolik, A. & Myers, S.L. (2024). How Telegram became a playground for criminals, extremists and terorists. *New York Times*, September 7, 2024. https://www.nytimes.com/2024/09/07/technology/telegram-crime-terrorism.html?smid=nytcore-ios-share&referringSource=articleShare&sgrp=c-cb.

National Democratic Institute (2022). *Interventions for ending online violence against women and girls*. https://www.ohchr.org/sites/default/files/documents/issues/expression/cfis/gender-justice/subm-a78288-gendered-disinformation-cso-ndi-annex-3.pdf.

Norman, A. (2021). *Mental immunity*. New York: Harper.

North Atlantic Treaty Organization. (2023). *NATO's approach to countering disinformation*. https://www.nato.int/cps/en/natohq/topics_219728.htm.

Off, C. (2024). *At a loss for words: Conversation in the age of rage*. Toronto: Random House Canada.

Pain, P. (2023). "Suddenly we were the story": Women journalists, the #MeToo movement and online misogyny in India. In L.M. Cuklanz (Ed.), *Gender violence, social media and online environments* (pp. 113-129). London: Routledge.

Parliament of Canada, House of Commons Standing Committee on Public Safety and National Security. (2024). *Minutes of Proceedings.* 44th Parliament, 1st session, meeting no. 121. Retrieved from the Parliament of Canada website: https://www.ourcommons.ca/documentviewer/en/44-1/SECU/meeting-121/evidence.

Pomerantzev, P. (2024). *How to win an information war*. London: Faber.

Powell, A., & Sugiura, L. (2018). Resisting Rape Culture in Digital Society. In W. S. DeKeseredy, C. M. Rennison, & A. K. Hall-Sanchez (Eds.), *The Routledge International Handbook of Violence Studies* (pp. 469–479). Milton: Routledge.

Pronk, N.P., Hernandez, L.M., Lawrence, R.S. (2013). An integrated framework for assessing the value of community-based prevention: A report of the Institute of Medicine. *Prevention of Chronic Disease*,10:120323. DOI: http://dx.doi.org/10.5888/pcd10.120323.

Public Inquiry Into Foreign Interference in Federal Electoral Processes and Democratic Institutions (2025). *Final report. Volume 1: Report summary*. His Majesty the King in Right of Canada. https://foreigninterferencecommission.ca/fileadmin/report_volume_1.pdf.

Ressa, M. (2022). *How to stand up to a dictator*. New York, NY: Harper Collins.

Ribeiro, M. H., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., Long, S., Greenberg, S., & Zannettou, S. (2021). The evolution of the manosphere across the web. In *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media* (pp. 196–207). AAAI Press. https://ojs.aaai.org/index.php/ICWSM/article/view/18053/17856.

Richardson-Self, L. (2021). *Hate speech against women online: Concepts and countermeasures*. Lanham, MD: Rowman and Littlefield.

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. New York: Picador.

Roozenbeek, J., van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*(65). https://doi.org/10.1057/s41599-019-0279-9.

Ryan, J. (2025). Europe's race to arm is pointless if its adversaries are waging war online. *The Guardian*, April 15, 2025. https://www.theguardian.com/commentisfree/2025/apr/15/us-europe-military-spending-trump-ireland?CMP=Share_iOSApp_Other.

Samson, D.R. (2023). Out tribal future: How to channel our foundational human instincts into a force for good. New York, NY: St. Martin's Press.

Schick, N. (2020). *Deepfakes: The coming infocalypse*. New York, NY: Twelve.

Secrétariat général de la défense et de la sécurité nationale (VIGINUM) (2024). Matryoshka: A pro-Russian campaign targeting media and the fact-checking community. https://www.sgdsn.gouv.fr/files/files/20240611_NP_SGDSN_VIGINUM_Matriochka_EN_VF.pdf.

Sugiura, L. & Smith, A. (2020). Victim Blaming, Responsibilization and Resilience in Online Sexual Abuse and Harassment. In: Tapley, J., Davies, P. (eds) *Victimology*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-42288-2_3.

Sobieraj, S. (2020). *Credible threat: Attacks against women online and the future of democracy*. Oxford, UK: Oxford University Press.

Sorell, T. & Kelsall, J. (2025). Violent video games, recruitment and extremism. *Criminal Justice Ethics*. DOI: 10.1080/0731129X.2025.2484974.

Springer. F. & Phillips, J. (n.d.). *The institute of medicine framework and its implication for the advancement of prevention policy, programs and practice*. Center for Substance Abuse Prevention. http://ca-sdfsc.org/docs/resources/SDFSC_IOM_Policy.pdf

Stanley, J. (2024). *Erasing history: How fascists re-write the past to control the future*. New York: Simon and Schuster.

Stark, E. (2007). *Coercive control: How men entrap women in personal life*. Oxford University Press.

Stuart, K. (2025). Video games can't escape their role in the radicalisation of young men. *The Guardian*, March 24, 2025. https://www.theguardian.com/games/2025/mar/24/video-games-cant-escape-their-role-in-the-radicalisation-of-young-men?CMP=Share_iOSApp_Other.

Susmann, M.W. & Wegener, D.T. (2022). The role of discomfort in the continued influence effect of misinformation. *Memory and Cognition*, 50, 435-448. https://doi.org/10.3758/s13421-021-01232-8.

Tenove, C., Tworek, H.J.S., & McKelvey, F. (2018). *Poisoning Democracy: How Canada Can Address Harmful Speech Online*. Public Policy Forum. https://ppforum.ca/wp-content/uploads/2018/11/PoisoningDemocracy-PPF-1.pdf.

Thakur, D. & Hankerson, D.L. (2021). *Facts and their discontents: A research agenda for online disinformation, race, and gender.* Center for Democracy & Technology. https://osf.io/preprints/osf/3e8s5.

UN Women Expert Group (2022). *Technology-facilitated violence against women: Towards a common definition. Report of the meeting of the Expert Group*. World Health Organization. https://www.unwomen.org/sites/default/files/2023-03/Expert-Group-Meeting-report-Technology-facilitated-violence-against-women-en.pdf

van der Linden, S. (2013). 'What a hoa'. *Scientific American Mind*, 24(4), 40-43.

van der Linden, S. (2015). The conspiracy effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences, 87*, 171-173.

van der Linden, S. (2021). The best way to deal with Covid myths this Christmas? Pre-bunk rather than debunk. *The Guardian*, December 23, 2021. https://www.theguardian.com/commentisfree/2021/dec/23/covid-myths-christmas-vaccines-virus-missinformation.

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 28 (March), 460-467. DOI: 10.1038/s41591-022-01713-6.

van der Linden, S. (2023). *Foolproof: Why misinformation infects oru minds and how to build immunity*. London, UK: Norton.

van der Linden, S., Panagopoulos, C., & Roozenbeek, J. (2020). You are fake news: Political bias in perceptions of fake news. *Media, Culture & Society, 42*(3), 460–470. https://doi.org/10.1177/0163443720906992.

Whyte, C. (2020). Cyber conflict or democracy "hacked"? How cyber operations enhance information warfare, *Journal of Cybersecurity*, *6(1)* https://doi.org/10.1093/cybsec/tyaa013.

Zmigrod, L. (2022). A psychology of ideology: Unpacking the psychological structure of ideological thinking. *Perspectives on Psychological Science, 17*(4), 1072-1092. DOI: 10.1177/17456916211044140.

Zmigrod, L., Burnell, R. & Hemeleers, M. (2023). The misinformation receptivity framework. *European Psychologist, 28*(3), 173-188. https://doi.org/10.1027/1016-9040/a000498.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New Yok, NY: Public Affairs.

ANNEXES

Annex A: Project Team

Community Safety Knowledge Alliance

Dr. Janos Botschner, PhD – *Project Lead.* Janos is a social scientist with deep experience in applied research and evaluation and strategic consulting across a range of contexts. He holds a joint doctorate in applied social and developmental psychology. Janos has held a number of adjunct faculty appointments and administrative positions during a lengthy career in the broader public sector. Janos' professional work covers the spectrum of issues related to collaborative public safety and community well-being, with a focus on understanding, and responding adaptively to, the complex issues and emerging opportunities of today's world.

Cal Corley, MBA Cal is CEO of the Community Safety Knowledge Alliance and a former Assistant Commissioner of the RCMP. Over the course of his career, Cal gained extensive experience in both operations and executive management, serving in such areas as national security, criminal intelligence, drug enforcement, human resources, and leading reform initiatives. He also served on secondments at the Privy Council Office and at Public Safety Canada.

Ritesh Kotak, JD, MBA is a Technology and Cybersecurity analyst and a licensed lawyer in Ontario. Ritesh started his career in public safety working for two police organizations focusing on cybercrime investigations and innovation. He left policing to pursue an MBA and then worked in Big Tech for two years focusing on innovation and smart cities. He left the Tech sector to attend law school and received a JD with a Law and Technology Option. Ritesh is a frequent contributor on mainstream media and is an international public speaker. Ritesh has also appeared twice as a witness in House of Commons Committees.

Sapper Labs Group

Dave McMahon, Hon. B.Eng., M.S.M. – *Project Co-Lead.* Chief Intelligence Officer at SLG, Dave is a deep generalist and expert with 40 years of experience in intelligence operations, cyber and cognitive warfare. He has an honours degree in Computer Engineering from the Royal Military College of Canada. Dave served with the Canadian Armed Forces, the Canadian Security Intelligence Service (CSIS), the Communications Security Establishment (CSE), the Security Intelligence Review Committee (SIRC), and the Office of the Communications Security Establishment Commissioner (OCSEC). He was a principal architect of a number of national offensive cyber and foreign intelligence programs for Canada. Dave co-chaired the interdepartmental committee on Information Warfare and psychological operations.

Dr. Giovanna Cioffi, CD, Hon. BA, MDEM, MES, PhD, served as an Army intelligence analyst r and expert in Cyber warfare and Psychological Operations. She was Deputy Chief of Targets/Ground Force Analyst/Special Purpose Reconnaissance Analyst (Operation IMPACT), a Captured Equipment and Material Analyst (Digital Forensics) (Op IMPACT), and a National Security Team Open Source Intelligence Analyst with a multinational joint intelligence task force covering global extremism.. She also worked as an Intelligence Operator and Intelligence Analyst at CANSOFCOM as well as Civil Military Cooperation and Psychological Operations Analyst/Tactical Operator with the CAF.

Dr. Juliane Ollinger, PhD is a research scientist with a PhD in Microbiology (Cornell 2008) with a focus on infectious diseases. Julie brings her strong research background and critical analysis skills to the Sapper team and has contributed to intelligence investigations focused on Due diligence, National Security, Foreign Interference, Disinformation and Support to Defence Operations.

Bradley Sylvestre, MA is an analyst with Sapper Labs Group focused on open-source intelligence (OSINT) and strategic analysis. His research interests broadly encompass strategic competition, foreign interference, espionage, disinformation and deep fake research. Prior to joining Sapper Labs Group, Bradley worked as a strategic analyst with the Canadian Armed Forces and Department of National Defence. Within the force development enterprise, his efforts supported work to identify the necessary capabilities to enable and sustain the Canadian Armed Forces and missions through current intelligence. Bradley holds a MA in International Affairs from the Norman Paterson School of International Affairs (NPSIA), also located in Ottawa.

Actua

Actua and CSKA collaborated to produce resources tailored to parents, youth and educators, based on knowledge synthesized by CSKA and SLG. The following staff members led Actua's involvement in this work.

Mikayla Ellis, BA, Senior Manager, Outreach.

Janelle Fournier, PhD (ABD), Senior Manager, Education.

Abbey Ramdeo, MT, Manager, National Educator Learning Program.

Annex B: Advisory Committee

Michael (Mike) Doucet is a senior leader of portfolios focusing on public safety and technology. He served as executive director of the Security Intelligence Review Committee, now known as National Security and Intelligence Review Agency. He currently serves as Executive Director, Office of the CISO, at OPTIV, a cyber advisory and solutions company, providing strategic advice on cyber programs, technology and risk.

Jennifer Flanagan is the President and CEO of Actua, which has become Canada's largest STEM outreach organization. It represents a national network of 43 universities and colleges that engage youth, ages 6-26, in STEM learning experiences, and advancing equity, diversity and inclusion in STEM. Actua's activities annually reach 350,000 young people. In 2021, Jennifer was awarded in the Manulife Science and Technology category, which recognizes women in STEM roles who are challenging the status quo for knowledge and female empowerment.

Dr. Carmen Gill is a professor in the Department of Sociology at the University of New Brunswick. She works in partnership with police agencies in Canada. Her research focuses on police intervention in intimate partner violence (IPC), domestic homicide and treatment of perpetrators and victims through the criminal justice system. Carmen is currently leading a three-year national research project entitled: Coercive control, risk assessment and evidence of intimate partner violence: Police response in partnership with the Canadian Association of Chiefs of Police (CACP), the Canadian Police Knowledge Network (CPKN) and l'École nationale de police du Québec. Carmen was previously the leader of the Canadian observatory on the justice system response to intimate partner violence (2006-2016). She led the development of the national framework for collaborative police action on IPV with CACP.

Jennifer Irish has more than 20 years of experience in foreign service and diplomacy. She has been appointed to postings in various embassies across the world, and has been a part of Canada's Permanent Missions to the United Nations in Geneva and New York. In addition to serving three terms in Canada's Privy Council Office, she worked as Director General at Canada's Integrated Terrorism Assessment Centre. Jennifer currently works as an Associate at University of Ottawa's Telfer Centre for Executive Leadership, co-directing its Canada Security and Intelligence Leadership program and teaching accountability, critical thinking, and strategic communication to Canada's next generation of leaders.

Alan Jones is executive adviser to the University of Ottawa Professional Development Institute and a retired CSIS officer who served in numerous operational and policy positions, including assistant director of CSIS. Alan's CSIS career included being the Chair of the G8 working Committee on Terrorism, Senior Policy Advisor in the Privy Council Office, Security and Intelligence Secretariat, Director General of the Counter Terrorism Branch and Director General of the International Terrorism Branch. In 2008 Alan became the Assistant Director for Operations, responsible for all

operational programs and in 2010 he became the Assistant Director for Technology which included both corporate and operational technology.

Marcus Kolga is an international award-winning documentary filmmaker, journalist, digital communications strategist, and a leading Canadian expert on Russian and Central and Eastern European issues. Marcus has a focus on communications and media strategies as tools of foreign policy and defence, and continues to write commentary for national and international media including the Globe and Mail and Toronto Star. He is the co-founder and publisher of UpNorth.eu, an online magazine that features analysis and political and cultural news from the Nordic and Baltic region. Marcus is involved with international human rights organizations and national political organizations. In 2015, Marcus was awarded the Estonian Order of the White Star by President Toomas Hendrik Ilves.

Annex C: System of People, Processes and Technology Aligned to Theory of Change

Levels and outcomes from theory of change addressed by the present project

Focus of Interventions	Ecological Levels			
	Individual	**Microsystem**	**Exosystem**	**Macrosystem/ Chronosystem**
	Physical, mental & social development & wellbeing	Family, peers, schools, religious groups, health system	Neighbours, legal & social welfare services, community-based services, mass & social media	Attitudes & ideologies of broader culture/ Major global & environmental events occurring over time
Society			Acts of misogyny are less tolerated GD-based foreign interference is detected and identified	
Government			Greater accountability for enablers & perpetrators of GD GD-based foreign interference is detected and identified Foreign interference/transnational oppression is proactively targeted	
Population			Public awareness of GD-based foreign interference	
Organizations & Networks		Organizations & individuals use knowledge and techniques to identify & counter GD	Enhanced systems of support for victims of GD GD-based foreign interference is publicly identified	Greater public awareness about GD, its methods & its impacts
Individuals	Organizations & individuals use knowledge and techniques to identify & counter GD			

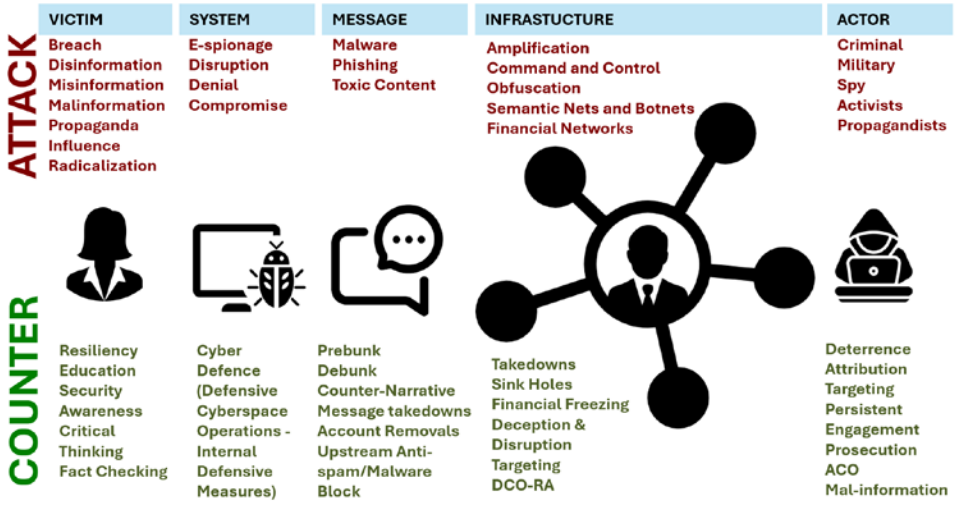
The main focus of the present project

Project outputs, corresponding ecological levels and areas of primary focus

Focus of Interventions	Ecological Levels			
	Individual	**Microsystem**	**Exosystem**	**Macrosystem/ Chronosystem**
	Physical, mental & social development & wellbeing	Family, peers, schools, religious groups, health system	Neighbours, legal & social welfare services, community-based services, mass & social media	Attitudes & ideologies of broader culture Major global & environmental events occurring over time
Society				
Government				**Knowledge products** Information sheet for government re GD as a national security threat/avenues for contending with GD in cases of foreign interference/transnat'l oppr'n **Options for addressing GD via policy, standards, regulations** Information sheet for organizations & government (consistent with theory of change)
Population		**Awareness & action focused knowledge products** Information sheets for individuals		
Organizations & Networks		**Awareness & action focused knowledge products** Information sheets for individuals & organizations **Additional resources** Literature/research synthesis	**Awareness & action focused knowledge products** Information sheets **Solution system** Technical options and framework for application of countermeasures – people, processes & technologies/tools) to counter GD and enhance system capacity **Recommendations for organizations and networks supporting those impacted by GD** **Engagement of key actors for KT/E**	**Options for addressing GD via policy, standards, regulations** Information sheet for organizations & government (consistent with theory of change) **Engagement of key actors** For KT/E and to enhance capacity and resilience
Individuals	**Awareness & action focused knowledge products** Information sheets for individuals			

Proposed system of people, processes and technology for countering gendered disinformation, aligned to theory of change

Cyber disinformation ecosystem and targeted countermeasures



VICTIM/AUDIENCE

- Building **resiliency** within the target audience and users starts with **security awareness** education, **critical thinking** and promoting access to **authoritative sources** of information.
- **Fact-checking**, media **literacy** programs and increased **transparency** in social media advertising can help the audience make informed decisions.
- **Cyber safe** programs and access to trusted end-device apps, platform and upstream security services.
- **Counter conspiracy beliefs** without challenging a **person's identity** may therefore be an effective strategy.

MESSAGE/PAYLOAD

- We can tackle **toxic messaging** and **malware** with content-based spam and malware filters supported by artificial intelligence (AI), **debunking/pre-bunking** of posts and **suspending accounts** of maligned influencers.
- **Counter-narratives** are highly-effective but should always be **truth-based** as part of an **information peacekeeping** strategy (IPK) or global peace and stabilization operations.

INFRASTRUCTURE

- Disinformation campaigns rely on cyberspace to propagate and amplify their malicious content or message with **botnets**. Cyberspace also offers an effective means to hide through **obfuscation** and non-attribution networks.
- **Enumerating** foreign global disinformation infrastructure requires effective **open-source intelligence** and targeting resources.
- Protective DNS services like **Canadian Shield** help.
- Ultimately **taking down** or **sink-holing** malevolent infrastructures has been a more **effective strategy** than chasing billions of toxic messages consumed by a target audience.

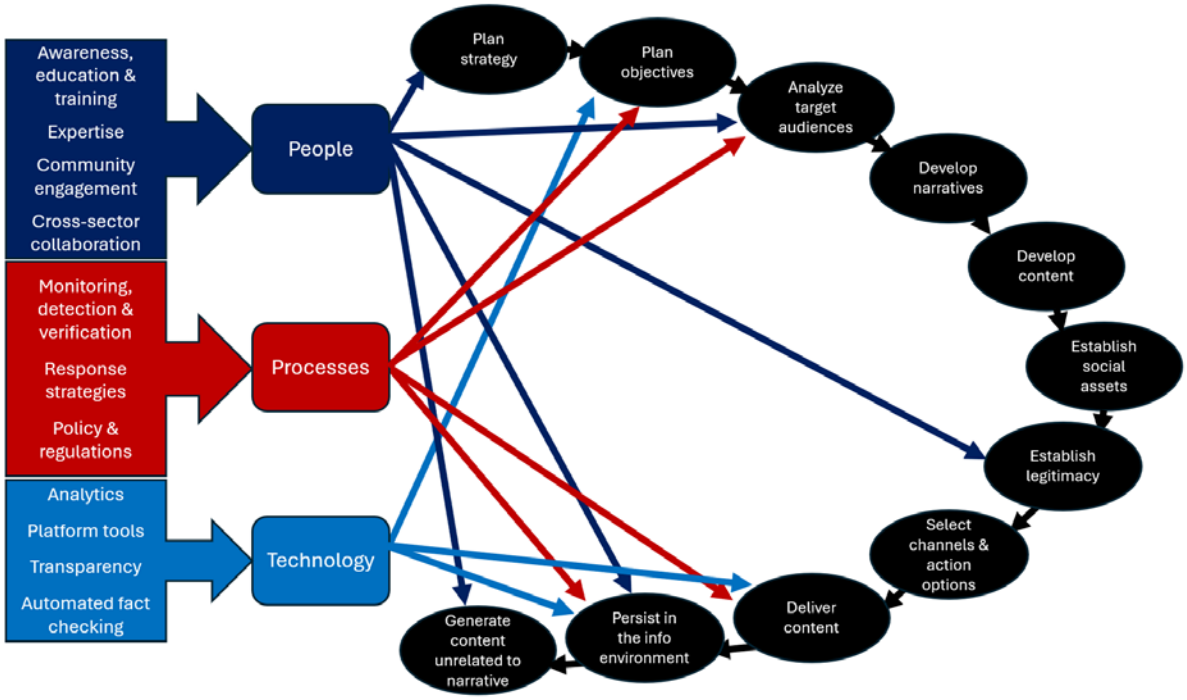
ACTOR

- The threat **actor** sits at the top of the food chain. Whether that is a **hostile intelligence service** (HoIS) or **paramilitary**or **transnational criminal organization**, **troll** farms or person-of-influence.
- Targeting the threat actor requires **strong attribution** substantiated with sophisticated **intelligence**, but is worth the effort.
- Cut the head off the troll and the disinformation campaign goes silent. Effects can include **sanctioning** companies and individuals, **freezing assets**, dismantling **financial networks**, **disrupting command and control**, maintaining **persistent engagement**, or following through with **indictment** and **prosecution**.
- **Industry** has been actively disrupting and **dismantling adversary networks**, **exposing** and **prosecuting actors** effectively for quite some time. **Public-private** collaboration, where appropriate, will be important from the perspectives of: effectiveness, accountability and trust.

PERSISTENT ENGAGEMENT

- The importance of **persistent engagement** at its core is to preserve our advantages and defend national interests in, through and from cyberspace by **contesting adversaries'** malicious cyber and influence activity during **day-to-day competition**.
- Strategic advantage is achieved through operations that **hunt** down the threat, **close the attribute chain**, **defend forward**, **contest** and **counter** the adversary **in real-time**.

Influence operation kill chain components addressed by proposed system of people, processses and technology



System features		Influence operation kill chain components targeted by proposed system of PPT										
		Plan strategy	Plan objectives	Analyze target audience	Develop narratives	Develop content	Establish social assets	Establish legitimacy	Select channels & action options	Deliver content	Persist in the information environment	Generate content unrelated to narrative
People	Awareness, education & training	■		■				■			■	■
	Expertise	■		■				■			■	■
	Community engagement	■		■				■			■	■
	Cross-sector collaboration	■		■				■			■	■
Processes	Monitoring, detection & verification		■	■						■	■	
	Response strategies		■	■						■	■	
	Policy & regulations		■	■						■	■	
Technology	Analytics		■							■	■	
	Platform tools		■							■	■	
	Transparency		■							■	■	
	Automated fact checking		■							■	■	

System features		System focus re social ecosystem (engagement & knowledge mobilization)				
		Individuals	Orgzn's & Netwks	Population	Government	Society
People	Awareness, education & training	Knowledge products to promote awareness & help identify forms & harms of GD & apply critical thinking				
	Expertise		Engagement of key actors for KT/E Recommendations for organizations and networks supporting those impacted by GD		Engagement of key actors for KT/E Avenues for contending with GD in cases of foreign interference	
	Community engagement					
	Cross-sector collaboration					
Processes	Monitoring, detection & verification		Options for addressing GD via policy, standards, regulations, guidelines, including perpetrator accountability/ rehabilitation		Options for addressing GD via policy, standards, regulations, guidelines, including perpetrator accountability/ rehabilitation Avenues for contending with GD in cases of foreign interference	
	Response strategies					
	Policy & regulations					
Technology	Analytics	Companion resources for understanding contexts, causes, threats (actors, methods/tools), enablers and consequences (emphasizing role of AI based tools as key enablers)	Companion resources for understanding contexts, causes, threats (actors, methods/tools), enablers and consequences (emphasizing role of AI based tools as key enablers) Technical options for countermeasures		Companion resources for understanding contexts, causes, threats (actors, methods/tools), enablers and consequences (emphasizing role of AI based tools as key enablers) Avenues for contending with GD in cases of foreign interference	
	Platform tools					
	Transparency					
	Automated fact checking					

System features		Influence operation kill chain components targeted by proposed system of PPT				
		Plan strategy	Analyze target audience	Establish legitimacy	Persist in the information environment	Generate content unrelated to narrative
People	Awareness, education & training (Create and diffuse knowledge to support ...)	Awareness about broader agendas and uses of online platforms to promote GD/MDM	Understanding of the ways that MDM/GD campaigns seek to: exploit and exacerbate social divisions; build on existing conspiracy theories; & leverage social media vulnerabilities/ properties to deepen and amplify impacts	Critical thinking about the ways that media platforms, accounts and influencers may be compromised and/or inauthentic (fake) purveyors of GD/MDM	Tools (concepts, systems) to detect information assets, operational activities and other TTPs used by those who are conducting influence operations	Individual and collective ability to observe and discern attempts to obscure the presence of malicious exploits within the information ecosystem
	Expertise (Translate & share knowledge about...)	Translate knowledge about broader context of misogyny and the ways it interacts to enable and benefit from GD/MDM			Awareness of the TTPs used by malicious actors to conceal information assets and operational activity	
	Community engagement (Engage relevant stakeholders to understand, amplify knowledge & collaborate)	Foster awareness about broader reasons online platforms are used to promote GD/MDM				Individual and collective capacity to distinguish GD/MDM content from distracting 'noise' and direct counter measures toward the GD/MDM within the information ecosystem
	Cross-sector collaboration (Support coordinated, joined-up action to...)	Diffusion of knowledge of concepts and harms to enable shared understanding and joined-up action to prevent and address harms	Promote awareness of, pre-bunk/de-bunk: attempts at manipulating social divisions & conspiracy theories Provide inputs on regulatory/ voluntary options to moderate vulnerabilities of tech. platforms		Foster networked capacity to detect relevant signals from contrived noise & to collaborate on measures to locate, identify and counteract GD/MDM content within operations that include aspects of concealment, diversion and distraction	

System features		Influence operation kill chain components targeted by proposed system of PPT			
		Plan objectives	Analyze target audience	Deliver content	Persist in the information environment
Processes	Monitoring, detection & verification (Describe/ explore approaches to...)		Examine and promote awareness of platform vulnerabilities related to hosting and propagating malicious information content Promote awareness of the general features of conspiracy theories to enable the discovery of information 'viruses' Promote awareness of the general features of common forms of GD/MDM	Document verified threat intelligence on GD/MDM Explore options for sharing threat intelligence in support of networked capacity to identify and respond to information operations involving GD/MDM	
	Response strategies (Identify the features of promising counter-narratives/ counter-measures)		Build knowledge of the features of counter-narratives that may counteract conspiracy theories and other forms of GD/MDM	Explore options for developing and enabling networked capacity to identify and respond to information operations involving GD/MDM, based on shared threat intelligence, and updated policy and regulation (where relevant & appropriate)	
	Policy & regulations (Examine & consider objectives, principles and tools...)		Engage policy advocates and policy professionals on dialogue about the role of policies and regulations, and technical options for achieving an effective balance between control versus freedom of speech in policies and regulations targeting GD/MDM, including the development and promotion of false content including conspiracy theories, the use of concealment to evade detection, and the use of GD/MDM as part of foreign interference		
		Consider when, how, by whom and under which powers and authorities, identified information operations involving GD warrant the use of state resources to disrupt/degrade/defend forward against foreign adversaries			

System features		Influence operation kill chain components targeted by proposed system of PPT		
		Plan objectives	Deliver content	Persist in the information environment
Technology	Analytics (Detect, identify, document & track...)	Identify contexts in which GD/MDM operations commonly take place, and the harms that they seek to perpetrate (e.g., harms against individuals, groups, society); consider objectives and the probability of GD/MDM operation being underway	Identify situations where it is likely that GD/MDM is being delivered, and the formats in which it this delivery might be occurring (e.g., deepfakes)	Distinguish GD/MDM content from distracting 'noise' and direct counter measures toward the GD/MDM
	Platform tools (Identify & limit the harms of...)	Consider actors likely to be responsible for using technology platform features and operational environments (e.g., telecos) to conduct certain GD/MDM activities (where evidence exists and can be collected, document links to specific actors)	Document and disseminate information about the features (design & business) of technology platforms (news media, social media, others) that enable malicious content to achieve depth, and scale of penetration against target audiences Document and disseminate verified information about actors using GD/MDM, including foreign interference (threat intelligence sharing)	
	Transparency (Reveal and promote awareness of...)			
	Automated fact checking (Identify and support accuracy...)	Identify and share information about frequent targets of malicious information exploits and the characteristics of the exploits	Use pre-bunking (when possible) and/or de-bunking (when necessary)	

Focus re awareness and prevention of victimization

Stakeholder	Awareness
Individuals (Directly-targeted individuals, parents, partners, allies)	• If this has happened to you, or someone you know, you/they are not alone • It is not "rare" or "isolated" problem; it is prevalent in Canada and beyond (include statistics) • It is not acceptable; in some cases, it may be illegal • Harms of GD/MDM for individuals and society (e.g., coercive control, emotional/psychological abuse, discrimination, polarization, intimidation/fear, disenfranchisement) • Forms of GD/MDM and their mechanisms of action (e.g., conspiracy theories, manipulated media/deepfakes) • How GD/MDM is produced, distributed and consumed using the features and exploiting the vulnerabilities of technology platforms (e.g., design and structures of platforms that create psychological rewards, along with ease and speed, of 'likes' or re-posting • Identifying the necessary skills at each stage to counter GD/MDM
Organizations & networks (e.g., community-based agencies/NGOs, police, school boards/threat risk teams)	• GD/MDM is a form of technology facilitated violence against women and girls • It is not "rare" or "isolated" problem; it is prevalent in Canada and beyond (include statistics) • It is not acceptable; in some cases, it may be illegal • It can be part of coercive control or intimate partner violence designed to harass, threaten or intimidate those who are targets of this behaviour • It is part of a broader context of harm that oppresses and subjugates females to a male-dominated social hierarchy • It can contribute to online and physical environments that are unwelcome and/or unsafe for women and girls • In extreme cases, it may be understood as an attempt to incite similar behaviour by others, and/or physical harms against women • It can discourage the participation of women and girls in the life of Canadian society, including in positions of influence and leadership • How GD/MDM is propagated using the features and exploiting the vulnerabilities of technology platforms • Incorporate media literacy programming: Develop curricula to teach critical thinking skills and how to identify manipulated media, especially targeting women and marginalized groups (e.g., IREX's Learn to Discern (L2D) initiative builds communities' resilience against disinformation and hate speech) • Digital safety training: Provide guidance on protecting personal data and images online to prevent their use in deepfake creation. This includes establishing institutional protocols for reporting and responding to online attacks.

Stakeholder	Awareness
Organizations & networks (e.g., community-based agencies/NGOs, police, school boards/threat risk teams)	• Propose: review of current regulatory and legislative conditions that may help or hinder GD/MDM; and study of options for legal frameworks that would support enhanced action (e.g., content moderation, technical properties) by platforms, deterrence of would-be perpetrators, and accountability for both platforms and individuals • User empowerment: Engage platforms (or alternative avenues) such that users are provided with tools to control their online experience and report harmful content • Engage civil society, levels of government and political parties on creation of codes of conduct or declarations of principles for electoral periods that address gendered disinformation • Establish complaints referral and adjudication processes for gendered disinformation cases • Coordinate with social media platforms to enhance dissemination of credible information and restrict problematic content • Community based rumour management has been used by the Sentinel Project to lessen the risk of mass atrocities – attention to the role of various communities in amplifying or dampening GD may be an important component of a holistic response involving multi-stakeholder dialogue and collaboration, along with developing coordinated response networks where the risk of GD/MDM may be elevated
Government (Policy, legislation, direct action through authorized agencies)	

Focus re response capacity

Stakeholder	Capacity
Individuals (Directly-targeted individuals, parents, partners, allies)	• Identification: Critical thinking; safe use practices • Resistance: Critical thinking; fact checking; counter narrative skills, focusing on the message (ignore, evade, address – pre-bunk/de-bunk); Cambridge University's gamified training tools, such as Bad News, can help adults and youth develop skills for identifying MDM • Reporting: Organizations; law enforcement • Resilience: Dialogue and information sharing; peer support with resistance, managing impacts; formal services (e.g., women's support organizations, shelters)
Organizations & networks (e.g., community-based agencies/NGOs, police, school boards/threat risk teams)	• Prevention/risk reduction: Awareness training for employees and the people they serve; security policies and procedures (e.g., for employees, service users, students) • Identification: Technological and non-technological tools, depending on technical maturity of the organization and its mandate and authorities • Response (e.g., counter measures): Counter narrative skills, focusing on the message (ignore, evade, address – pre-bunk/de-bunk) - e.g., develop and deploy counter speech campaigns (e.g., correct, de-emphasize false gendered content) to combat gendered disinformation • Reporting: Dialogue, information sharing, collaboration, MOUs/service protocols • Explore multi-sector partnering with civil society organizations to build coalitions to enhance monitoring capabilities

Stakeholder	Capacity

Government (Policy, legislation, direct action through authorized agencies)	Regulatory and legislative options, implications, opportunities and challenges
	
	• Convening, enabling and leadership:
	  ○ Establish a national task force to study and address gendered disinformation.
	  ○ Fund creation of a consortium of NGO and academic partners to conduct research, and serve as a clearing house for information, on: the nature, prevalence and impacts of GD/MDM (independent monitoring); inoculation strategies; technological innovations; and citizen literacy programs focusing on women and girls; public policy and regulatory options; technology governance; and legal/freedom of speech issues. Use national task force to govern creation and implementation of strategic information agenda and accompanying KT/E plan.
	  ○ Create a cross-departmental committee having external representation from national task force and consortium to prioritize attention to gendered disinformation: in foreign policy; and in processes shaping technology and disinformation policy.
	  ○ Explore opportunities to enhance coordination between platforms, fact-checkers, and election authorities
	
	Legal, enforcement options
	
	• Explore and assess legal protections that may be enacted to criminalize the creation and distribution of non-consensual pornography and malicious deepfake content.

Stakeholder	Capacity

Government (Intelligence and direct action supporting national security - e.g., through prevention/response to foreign interference and transnational oppression)
 	
(Potential for collaboration with private sector organizations, where appropriate and consistent with relevant legislative authorities)
	Cyber defence capacity (strategies, actions & techniques) taken by state of organizations to protect information ecosystems from cyber threats
	
	• Defensive cyberspace operations (DCO) – broad, strategic approach:
	  ○ Proactive measures: Prevention; proactive deterrence/disruption at-source; threat identification and intelligence sharing; cross-sectoral cooperation; international cooperations and cyber diplomacy to identify, track and respond to cyber threats
	  ▪ Target actors and their infrastructures through threat intelligence, targeting, takedowns, disruption, deception, prosecution etc – by integrating the F3EAD targeting framework (Find, Fix, Finish, Exploit, Analyze and Disseminate) and the DISARM framework (Detect, Interpret, Segment, Analyze, Respond, and Mitigate) and applying these to the Influence Operation Kill Chain (Annex A)
	  ○ Reactive measures (e.g., incident response) to protect the cyber domain from a range of cyber aggressions
	
	• Internal defensive measures (IDM) within organizations or networks
	  ○ Attention to reducing vulnerabilities and strengthening systems against influence operations involving GD/MDM

Focus re organizations and networks supporting those impacted by gendered disinformation

Stakeholder	Awareness	Capacity
Organizations & networks service providers (e.g., community-based agencies/NGOs, police, school boards/threat risk teams)	Information sheet supporting awareness of GD/MDM for stimulating awareness of GD/MDM & outlining	• Prevention/risk reduction: Awareness training for employees and the people they serve; security policies and procedures (e.g., for employees, service users, students); information on potential impacts for victims of GD/MDM • Identification: Awareness training focusing on GD/MDM as technology facilitated violence against women which may, in some cases, be an instance of coercive control/intimate partner violence • Response (e.g., counter measures): Identification and availability of legal and support resources for supporting victims of violence against women towards those impacted by GD/MDM. Develop and implement institutional protocols to support those attacked and address reports of attacks. Consider and assess opportunities for collaborative approaches to providing support services for targets of broadly focused gendered disinformation campaigns. • Reporting: Dialogue, information sharing, collaboration, MOUs/service protocols, exploration of policy and legal options with police, crowns, policy actors, women's groups and others

Annex D: Curated Sample Technology Options for Individuals, Human Service (Including Police) and Educational Organizations

ANALYTICS/AUTOMATED DETECTION

Note: Weblinks have not been provided as these may change, over time. Subscription fees, where indicated are current, as of April, 2025.

1. **Sentinel Deepfake Detection System**
   - **What it does**: AI detection platform that works with governments, media, & defence agencies to protect democracies from disinformation campaigns, synthetic media & information operations.
   - **How to use it**: Users can report gendered deepfakes for review.
   - **Subscription**: No public access; used by governments, media, & defence agencies.
   - **Example**: A deepfake targeting a female journalist is flagged & removed before going viral.

2. **WeVerify, DuckDuckGoose, DeepfakeProof**
   - **What they do**: Content verification, tracking, & debunking (WeVerify); AI powered deepfake detection for images, videos, & audio (DuckDuckGoose); Helps users identify deepfakes while browsing the web (DeepfakeProof).
   - **How to use them**: Chrome Plugin (WeVerify); Upload files via a regular browser to DuckDuck Goose; As a real-time deepfake detection plugin for Chrome (DeepfakeProof).
   - **Subscription**: Free/Open source platform (WeVerify); Subscription Required (DuckDuckGoose); Free Chrome Plug-in (DeepfakeProof).
   - **Example**: A fake nude image of a female politician is detected & debunked.

3. **Reality Defender**
   - **What it does**: Equips enterprises, governments, & platforms with the tools to detect AI generated or manipulated content in real time.
   - **How to use it**: Upload content to the software for real-time video identity, image & text authentication.
   - **Subscription**: Subscription required.
   - **Example**: A fake video targeting a women's rights activist is debunked before being used in a smear campaign.

4. **MeVer: Verification, Media Analysis, & Retrieval**
   - **What it does**: Developing technologies & services for understanding, searching, & verifying media content
   - **How to use it**: Journalists & researchers analyze disinformation content & networks.

- **Subscription**: Offers resources (tools, software, & datasets) via GitHub & other repositories.
- **Example**: A smear campaign against female journalists is traced to coordinated disinformation actors.

5. **RAND's Countering Truth Decay Initiative**
   - **What it does**: RAND researchers are studying the causes, consequences, & means of countering truth decay.
   - **How to use it**: Free resource.
   - **Subscription**: Research available on RAND's website for free.
   - **Example**: A journalist or researcher may explore Truth Decay research & commentary to understand the drivers, trends, & consequences of Truth Decay as a System.

PLATFORM/CONTENT GENERATOR TOOLS

1. **SynthID (Digital Watermarking)**
   - **What it does**: Watermarks & identifies AI generated content by embedding digital watermarks directly into AI generated images, audio, text, or video.
   - **How to use it**: Integrated into AI-generated media, detected by compatible tools.
   - **Subscription**: Available via Google Cloud's AI tools (Google DeepMind).
   - **Example**: A fake image of a female CEO is debunked using SynthID detection.

TRANSPARENCY

1. **Hoaxy – Tracking Gendered Disinformation**
   - **What it can do**: Hoaxly visualizes the spread of information online using the X/Twitter & Bluesky APIs.
   - **How to use it**: An API is used to retrieve recent posts matching your search query.
   - **Subscription**: Free until Hoaxly reaches its monthly post limit, then live search is only available to users with Basic ($100/month), Pro ($5000/month), or Enterprise (price available upon request) access.
   - **Example**: Hoaxy reveals bot activity pushing a false claim against a female official.

2. **Systematic Data Collection & Reporting**
   - **What it can do**: Track trends in gendered disinformation & AI-generated attacks.
   - **How it can be used**: Governments, researchers, journalists, & civil society can utilise reports for situational awareness, policy development, & advocacy.
   - **Subscription**: Varies - there is a wide variety of open source reporting available.
   - **Example**: A media watchdog report documents rising deepfake attacks on female politicians, which provides a situational awareness on deepfake trends.

3. **Gender-Sensitive Monitoring**
   - **What it can do**: One can utilise AI tools (e.g. Reality Defender), social network analysis (Hoaxy, Never), & qualitative methods to track gendered disinformation.
   - **How it can be used**: Quantitative / qualitative research to identify gendered attacks online.
   - **Subscription**: Varies - some tools are free, others require paid access.
   - **Example**: Through gender-sensitive monitoring, a researcher is able to show that women candidates face twice as many disinformation attacks as men.

4. **Enhanced User Reporting for Harmful Content**
   - **What it can do**: Improve response time & categorization of gendered disinformation reports.
   - **How it can be used**: As a mass-reporting campaign.
   - **Subscription**: Unknown, would depend on platform implementation.
   - **Example**: A journalist targeted by deepfakes reports it to an enhanced moderation system.

5. **Global Coalition for Digital Safety (World Economic Forum)**
   - **What it can do**: Develop politics & global coordination on digital safety, including gendered disinformation.
   - **How it can be used**: Advocacy groups can engage with the coalition to push for stronger policies.
   - **Subscription**: Dependent on how the coalition is set up.
   - **Example**: An NGO joins the coalition to push for stricter deepfake detection on social media.

Annex E: List of Accompanying Knowledge Resources

E1: Tackling Online Gendered Disinformation: Educator Guide

E2a: Tackling Online Gendered Disinformation: A Family Resource

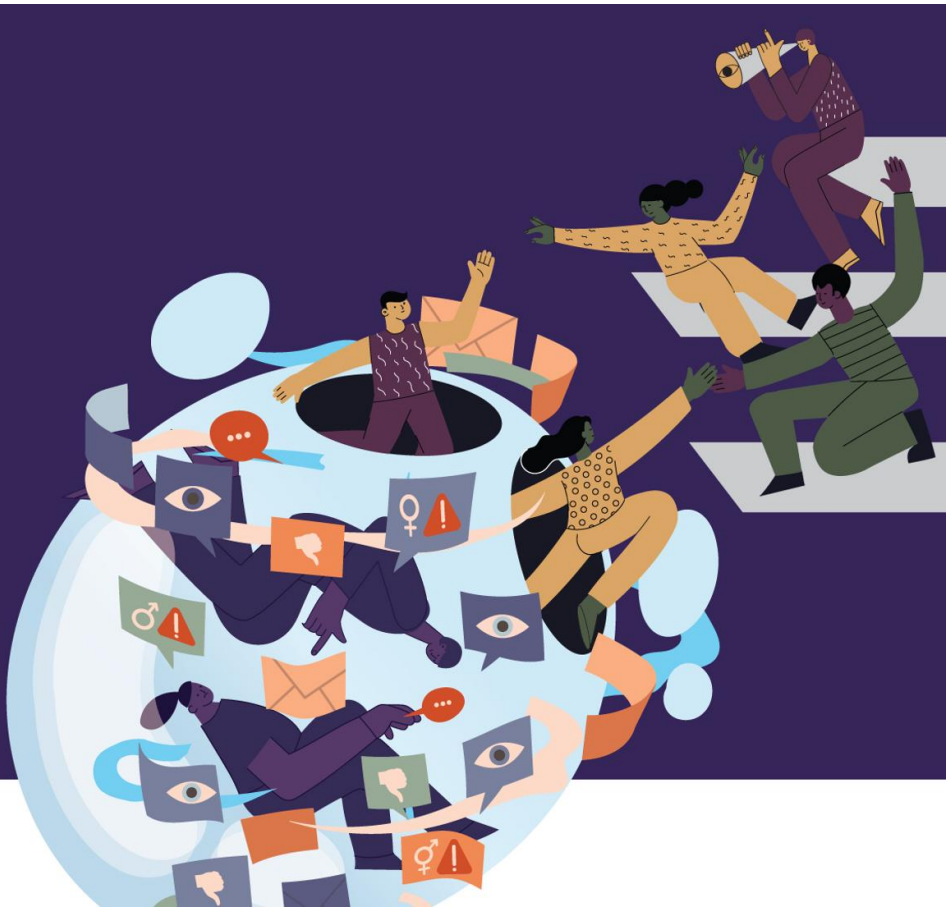
E2b: Tackling Online Gendered Disinformation: Youth Guide

E2c: Tackling Online Gendered Disinformation: Additional Resources for Educators, Families & Youth

E3: Gendered Disinformation: A Resource for Police and Human Service Agencies

E4: Knowledge Resources for Government

Understanding and Countering Gendered Disinformation



Annex E1

Tackling Online Gendered Disinformation: Educator Guide

CSKA Community Safety Knowledge Alliance

SAPPER LABS

DEVELOPED IN APRIL 2025

Tackling Online Gendered Disinformation

Educator Guide

actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

Table of Contents

Table of Contents

About Actua

Actua is creating a Canada where every child has the skills and confidence they need to achieve their full potential. As a leading science, technology, engineering and mathematics (STEM) outreach organization, Actua includes over 40 universities and colleges, engaging 500,000 youth in 600 communities each year. For 25 years, Actua has focused on identifying and removing the barriers for entry into STEM and now have national programs dedicated to engaging Indigenous youth, girls and young women, Black youth, those facing economic barriers and youth in Northern and remote communities.

This work is the product of collaboration between **Actua** and the **Community Safety Knowledge Alliance**, with **Sapper Labs Group**, and was supported, in part, through funding from Heritage Canada.

The Community Safety Knowledge Alliance (CSKA) is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes. Over the past decade, CSKA has conducted interdisciplinary research and engaged with change-makers on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

Sapper Labs Group (SLG) conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network. The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

Background Information: What and why

The fast-changing nature of our digital environment presents complex challenges, one of the most serious being gendered disinformation. Defined as the use of digital tools to inflict harm (physical, sexual, psychological, social, political, or economic) or violate the rights, freedoms and credibility of women and gender-diverse individuals, gendered disinformation is more than just online negativity. These are often coordinated campaigns with severe repercussions for women and other individuals marginalized due to gender, as well as to the health of our society and democracy in general.

To help address these challenges, we have created new resources to support educators, families and young people. These tools are designed to build awareness, encourage critical thinking and offer strategies for navigating and responding to disinformation they encounter online.

For educators to effectively teach these concepts, it's essential to grasp the complexities of the topic. **The following section defines gendered disinformation, provides examples, and explores the underlying factors driving its spread.** While much of the focus is on women and girls, it's important to recognize that gendered disinformation also impacts non-binary, trans, and gender non-conforming individuals. The dynamics of disinformation targeting these groups may be shaped by similar gender-based attitudes and ideologies or be used deliberately to deepen social divisions and undermine the ability of people in Canada to connect and collaborate.

Following this overview, we provide concrete steps educators can take to address gendered disinformation in their classrooms and communities.

What is gendered disinformation?

A politician faces a wave of faked images and videos designed to damage her credibility as a leader along with online threats to intimidate her into silence.

A student is targeted for abuse by an angry ex-boyfriend who spreads non-consensual explicit images of her on social media.

A teenage girl dreams of becoming a journalist. She shares her opinions online, only to be bombarded with hateful messages and false accusations that she's spreading lies.

A non-binary student asks their teacher about pronouns, but their classmates have already been misled by online posts claiming that there is no science behind non-binary identities, and that they are "made up."

These are all examples of gendered disinformation — false or misleading information designed to harm people based on their gender. It can take many forms, from online harassment, controlling behaviours, and manipulated images to false narratives that undermine the credibility of women and gender-diverse individuals. And while false information spreads quickly for many reasons, when it targets gender, it becomes a powerful tool for reinforcing discrimination and silencing voices.

DISINFORMATION VS. MISINFORMATION

- **Misinformation** is untrue content that is spread by people who believe that it is true. Misinformation could be spread innocently, or to cause harm.
- **Disinformation** is untrue content that is spread by people who know that it is untrue. Disinformation is always spread knowingly and deliberately to cause harm.

IMPACTS OF SPREADING DISINFORMATION

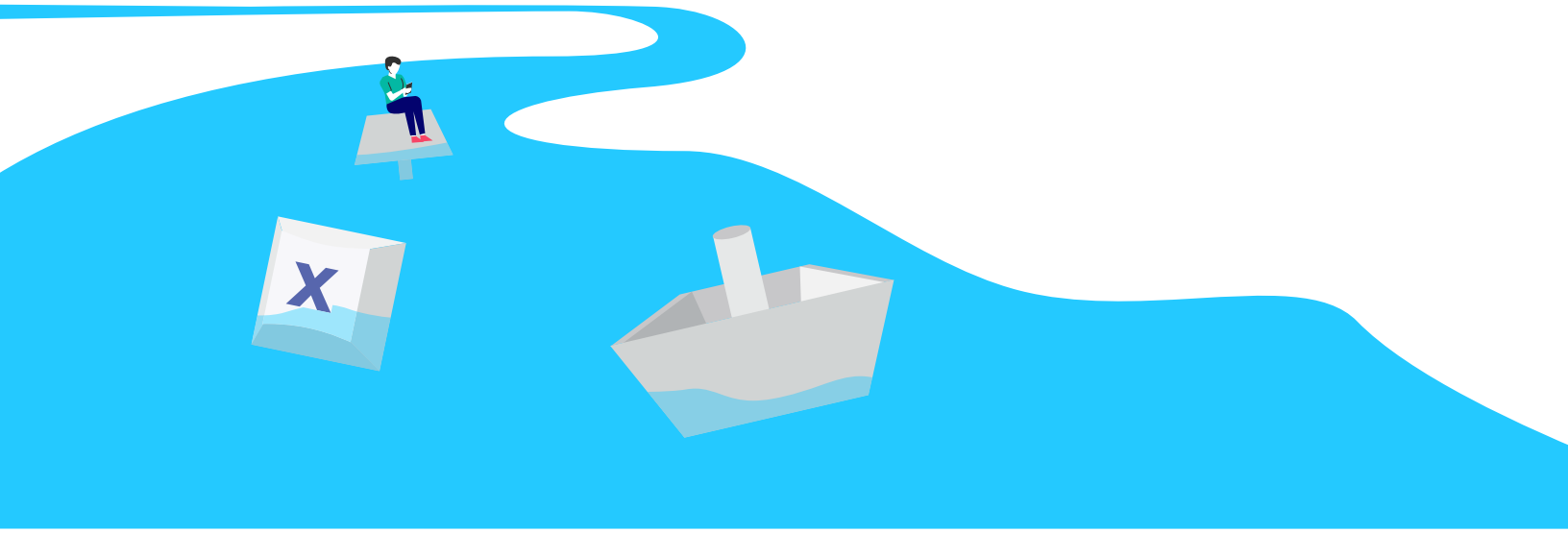
When these fake messages spread – usually through social media – they can make people question their abilities, limit their opportunities, or even fear speaking out online.

Seeing negative stereotypes over and over can cause self-doubt, anxiety, and low self esteem. You might start to question your abilities or feel pressured to fit into certain expectations. For example, if you constantly hear that "women aren't good leaders," you might be less likely to put yourself forward for leadership roles at school or in your community.

Recognizing gendered disinformation is the first step in countering its impact. Below are examples of how it appears in different contexts:

- **Fake stories** – Fake news articles or social media posts that attack women, especially those in leadership roles.
- **Manipulated images and videos** – Edited pictures or deepfake videos that make it look like someone said or did something they never did.
- **Misinformation about gender roles** – Posts or comments claiming that women are naturally bad at specific tasks or in certain sectors like leadership, science, or sports.
- **Harassment and cyberbullying** – Online attacks that try to intimidate, humiliate, or silence women and girls.
- **Memes and satire** – Jokes or cartoons that disguise harmful messages about women as "just humor."
- **Classrooms** – Comments or posts with the narrative that "girls are naturally less capable in math and science," discouraging female students from pursuing STEM careers.
- **Sports** – Women athletes often face public scrutiny amplified through social media, including shaming, objectification, sexist language, and debates over who is deemed eligible to compete in women's sports.

Can you think about other examples that you may have experienced or heard about?



Why is this happening?

The digital world is full of opportunities but also risks. Social media platforms reward outrage and engagement, meaning that harmful and false content often spreads faster than the truth.

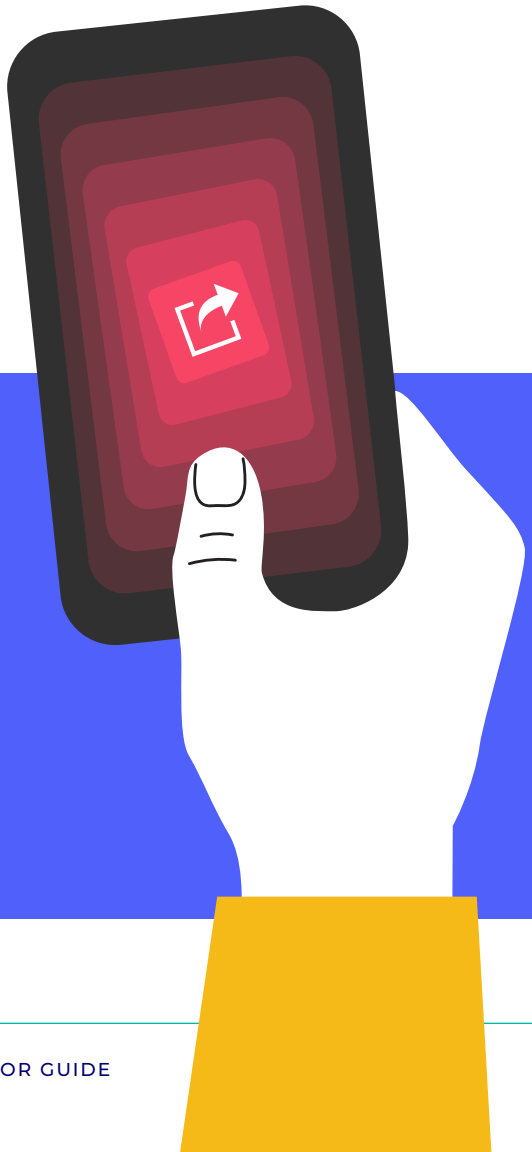
Certain groups — including women in public life and non-binary people — are especially vulnerable. Disinformation can be used to attack individuals or entire communities, making them feel unsafe or unwelcome in online spaces. Many individuals, especially young people, struggle to tell fact from fiction in the digital age. Algorithms push them toward content that reinforces their existing beliefs, creating "echo chambers" where misinformation flourishes. Some believe and share harmful content not because they want to hurt others, but because they don't realize they're being manipulated.

A FEW STATISTICS	SOURCE
• **1 in 5 Canadian women** experience some form of online harassment. • **30% of Indigenous women** encounter unwelcome behaviour online.	Canadian Women's Foundation (2024)
• Online abuse of women is a **widespread problem across all continents**.	Global Politics and the Vital Voices Global Partnership on TF-GBV (2024)
• Between 2019 and 2020, **85%** of women had **witnessed** or experienced online gender-based violence and 38% had been **personally impacted** by it • These figures likely **under-repor**t how widespread the issue is	Institute of Global Politics and the Vital Voices Global Partnership/Economist Intelligence Unit (2024)

TYPES OF TECHNOLOGY-FACILITATED ABUSE USED AS PART OF GENDERED DISINFORMATION

- **Doxxing** – When someone shares private information (like a home address or phone number) online to intimidate or harm a person.
- **Deepfake videos** – Fake videos created with artificial intelligence (AI) to make it look like someone is doing or saying something they never did. These are often used to spread false, damaging information about women.
- **Non-consensual image sharing** – When someone shares or threatens to share private photos without permission.
- **Fake accounts and impersonation** – Creating fake profiles to spread lies, harass someone, or damage their reputation.
- **Hate speech and threats** – Sending harmful messages or posting threats to silence women and girls online.

These forms of abuse can be intimidating, but identifying them is the first step in protecting yourself and supporting students in navigating online spaces safely.



Taking Action in Education

The good news is that we are not powerless against gendered disinformation. Tackling it requires a mix of education, community support, and technology — all working together.

Approaches in the school

Creating a school-wide culture that recognizes and addresses gendered disinformation requires collective action, strategic planning, and ongoing learning. Administrators play a key role in leading this shift by fostering an inclusive, supportive, socially and technologically aware environment that includes digital literacy and critical thinking:

- Provide opportunities for professional learning on digital citizenship and cyber safety, as well as digital literacy training for educators, including how to recognize and respond to gendered disinformation. Consider offering this training to all teachers across subject areas, as well as other school staff (e.g., coaches, educational assistants, administration), to support a shared foundation of awareness. Framing this as part of broader efforts to build capacity and strengthen organizational culture can help embed it as a key element of school-wide practice.
- Develop policies to address gendered disinformation and harassment within the school community, including reporting mechanisms and digital safety protocols.
- Encourage district-wide initiatives that promote inclusive digital citizenship.

Approaches in the classroom

Teachers are a key resource in equipping students with the skills to critically evaluate online content and challenge disinformation. Integrating these lessons into everyday classroom activities can build critical thinking, media literacy, and digital resilience.

Here are some ways we can take action:

1. TEACH PEOPLE HOW TO RECOGNIZE AND RESIST DISINFORMATION

We should be equipping students with the knowledge and tools to spot and explain fake news, AI-generated images, and other manipulative content. This means:

- Teaching how social media platforms (algorithms and other design features) work so they understand why they keep seeing the same kinds of posts and how posts can spread quickly.
- Showing how to check sources and modelling ways to think critically before sharing information.
- Encouraging students to share what they've learned with their friends and family members who are also vulnerable to disinformation and may not realize.

Check out the **"Entry Points and Subject Connections"** on *page 13* on how to integrate these conversations into the classroom.

2. USE EMPOWERING LANGUAGE OVER FEAR-BASED LANGUAGE

It is crucial to approach this discussion with empowering language so that youth are not afraid to be online. Yes, we want to make them aware of the risks, but by teaching them key skills we help them to build confidence to navigate online spaces in a smart and safe way.

- "The internet can be a good resource" vs. "The internet is scary"
- "It is important to ask the right questions" vs. "You need to be on guard and ready to protect yourself"

Follow up on this by empowering them to share what they've learned with their friends and family members so that they can extend their knowledge to others.

3. SUPPORT THOSE WHO ARE TARGETED

It's crucial to create safe, supportive spaces where students can talk about their online experiences and seek help when needed:

- Create safe spaces where people can talk about their experiences without fear.
- Talk about trusted networks (like teachers, youth workers, or family members) that can step in when someone is being harassed online.

Check out the **"Navigating Challenging Conversations Section"** on *page 17*.

4. PROMOTE INCLUSION, CRITICAL THINKING, AND WELLBEING

One way gendered disinformation spreads is when it is linked to conspiracy theories or movements that help people who have felt left out of society feel more powerful, at the expense of others. Supporting students to be resilient, flexible and critical thinkers can help. It is also important that students feel like school is somewhere they belong and where they can have a voice. Encouraging open, informed conversations, and helping people get the resources they need, may reduce susceptibility to polarization or radicalization.

This means:

- Encouraging young people to explore multiple identities – as learners, community members, and future leaders.
- Teaching empathy and critical thinking so people learn to question divisive narratives instead of falling for them.
- Creating inclusive spaces where gender diversity is understood and respected.
- Helping youth to build skills in communication and conflict resolution, especially with those that may have different lived experiences, perspectives or knowledge than them.

5. BUILD DIGITAL LITERACY INTO EVERYDAY LEARNING

Instead of waiting until problems arise, we should proactively teach digital literacy and ways to develop psychological immunity in schools and community programs. This means learning how to recognize and how to resist to mis- and disinformation:

- Introducing media literacy lessons as early as possible so young people grow up questioning what they encounter online. These lessons can include conversations to:
    - Advocate for better policies from tech companies to stop gendered disinformation from spreading, such as stronger protections and more transparent content moderation.
    - Help them become aware of how their data and engagement are used to fuel these systems – so they can make informed choices about their online activity.
- Seeking training to feel comfortable discussing mis- and disinformation, bias, and online safety.
- Using age-appropriate examples (such as AI deepfake videos or gendered memes) to show how disinformation works in real life.
- Talking about the features of mis- and disinformation (including conspiracy theories) and the social and information environments that make them attractive, "sticky" and hard to resist.

Below are some examples of **how** to bring these topics into your educational setting.

Opportunities in the Classroom

ENTRY POINTS AND SUBJECT CONNECTIONS

Here are some ideas for integrating this topic into your classrooms. Taking a cross-curricular approach helps reinforce key messages across different subjects, platforms, and school activities.

While we're sharing some STEM connections and entry points, these conversations don't have to be tied strictly to curriculum goals. Real-world connections (like a viral meme, news story, or events like Media Literacy Week) can provide natural opportunities to engage students in meaningful discussions.These forms of abuse can be intimidating, but identifying them is the first step in protecting yourself and supporting students in navigating online spaces safely.

As you explore these connections, think of what others you can make in your classroom.

SUBJECT	ENTRY POINTS AND SUBJECT CONNECTIONS	CROSS-CURRICULAR INTEGRATION EXAMPLES
STEM (Science, Technology, Engineering, Math)	**Scientific literacy and fact-checking:** Teach students how to assess scientific claims for credibility, including evaluating sources. Highlight how scientific mis- or disinformation spreads by misusing or misrepresenting scientific studies.	**Language connection:** Write research-based essays or opinion pieces evaluating misleading scientific claims (e.g., climate change).
	AI and scientific ethics: Examine how AI is used in scientific research and the ethical challenges of AI-generated mis- or disinformation.	**Ethics/philosophy connection:** Hold debates on AI ethics in scientific research, focusing on potential misuse of data.
	Data literacy: Teach students how to evaluate sources, analyze statistics, and recognize manipulated data.	**Math and language connection:** Write explanatory articles using real-world statistical data to debunk common mis- or disinformation.
	Math and statistics: Analyze how misleading statistics and manipulated data influence public perception.	**History connection:** Study how manipulated data has been used in historical propaganda (e.g., economic data during wars).
HISTORY AND SOCIAL STUDIES	**Propaganda in history:** Study the role of propaganda in shaping public opinion during key historical events (e.g., WWII, Cold War).	**STEM connection:** Explore how technology (e.g., early radio, AI-driven bots) has evolved to spread propaganda and mis- or disinformation.
	Civic engagement and misinformation: Analyze the impact of disinformation on democracy, elections, and trust.	**Language connection:** Have students write persuasive essays on how mis- or disinformation affects civic engagement today.
	Case studies: Compare historical and modern gendered mis- or disinformation.	**Media studies connection:** Analyze historical propaganda posters alongside modern digital memes spreading disinformation.

SUBJECT	ENTRY POINTS AND SUBJECT CONNECTIONS	CROSS-CURRICULAR INTEGRATION EXAMPLES
LANGUAGE AND MEDIA STUDIES	**Analyzing news and social media:** Compare media coverage of male and female politicians or athletes to uncover bias.	**STEM connection:** Investigate how AI algorithms amplify biased content in social media feeds.
	Fake news and deepfakes: Teach how AI-generated content, including images and videos, can spread disinformation.	**Computer science connection:** Code simple AI programs or simulations that mimic deepfake technology.
	Persuasive writing: Write fact-checked opinion pieces on the impact of disinformation.	**Psychology connection:** Incorporate cognitive bias studies to help students understand why people fall for fake news.
PSYCHOLOGY AND SOCIOLOGY	**Cognitive biases:** Teach how biases (e.g., confirmation bias) make people more vulnerable to believing disinformation.	**Language connection:** Analyze characters in literature who exhibit biases and how this influences their decisions.
	Language connection: Analyze characters in literature who exhibit biases and how this influences their decisions.	**Media studies connection:** Explore how news outlets use algorithms to decide headlines and content placement.
	Impact on identity: Discuss how gender stereotypes in digital spaces shape self-perception and confidence. Examine the ways that certain vulnerabilities may lead people to seek affiliations with groups that demand and reinforce strong "us/them" perspectives.	**Physical education connection:** Analyze how online stereotypes affect participation and leadership in sports. Examine the ways that seeing people as "other" can lead to them being excluded from teams or vilified as supporters of different teams.
PHYSICAL EDUCATION AND SPORTS	**Media representation in sports:** Compare how male and female athletes are portrayed in sports media.	**Language connection:** Write media critiques or sports columns analyzing gender bias in coverage.
	Gender bias in coaching and athletics: Discuss how stereotypes influence coaching and athletic opportunities.	**Psychology connection:** Explore how stereotype threat impacts athletic performance.
	Online harassment in sports: Examine cases of gendered disinformation targeting female athletes.	**Civics connection:** Discuss how online harassment reflects larger societal issues of gender equity and digital citizenship.

KEY CONSIDERATIONS

When teaching about gendered disinformation, it's important to tailor your approach based on students' age, developmental stage, and emotional readiness. Equally critical is being prepared to navigate difficult conversations, as these topics can sometimes provoke discomfort, resistance, or misunderstanding.

AGE APPROPRIATENESS

Understanding how to introduce complex topics at different ages helps foster engagement and comprehension. As students grow, the same topic can be explored in greater depth, strengthening their critical thinking, digital literacy, and awareness of gendered disinformation.
Check out Actua's suite of Cyber Smart activities for age appropriate activities on this topic from Gr. 2-12.

Below are examples of how topics related to gendered disinformation can be adapted and expanded across grade levels:

TOPIC	ELEMENTARY (GRADES 3-6)	MIDDLE SCHOOL (GRADES 7-9)	HIGH SCHOOL (GRADES 10-12)
SPOTTING FAKE CONTENT	**Basic recognition:** Show clearly edited images (e.g., unicorn in real life) and ask, "Does this look real? Why or why not?" Discuss why people might edit images or create fake stories.	**Analyzing misinformation:** Compare subtle examples, like misleading headlines or slightly altered images, and discuss their potential impact.	**Deep analysis:** Explore deepfake videos, AI-generated content, and viral mis- or disinformation. Discuss real-world consequences and how to fact-check effectively.
MEDIA BIAS	**Fact vs. opinion:** Teach students to differentiate between facts and opinions in simple news stories or social media posts.	**Bias in news sources:** Compare how different media outlets report the same event. Discuss how word choices or images might reveal bias.	**Bias, framing and algorithms:** Discuss how social media algorithms reinforce bias and shape the content users see. Explore ways to diversify their media consumption.

TOPIC	ELEMENTARY (GRADES 3-6)	MIDDLE SCHOOL (GRADES 7-9)	HIGH SCHOOL (GRADES 10-12)
IMPACT OF DISINFORMATION	**Story-based learning:** Read a story where a character spreads a false rumor and discuss the emotional impact on those involved.	**Emotional impact and online behaviour:** Discuss how false information and online harassment can affect people emotionally and socially.	**Social, ethical, and global impact:** Analyze real cases where gendered disinformation harmed individuals or communities and brainstorm ways to counter it.
CRITICAL THINKING AND DIGITAL RESILIENCE	**Basic critical thinking:** Encourage questions like, "Who made this? Why did they make it?" when evaluating digital content. Invite discussion of how a message might make us feel about a topic and how these feelings might be manipulated for negative purposes.	**Critical thinking and digital skills:** Teach students to fact-check, question sources, and recognize manipulative content. Conduct activities that develop self-awareness and the ability to notice how they react to the content of different messages.	**Advocacy and ethical responsibility:** Engage students in debates about digital ethics, free speech, and the responsibility of social media platforms in tackling disinformation. Explore the implications of these issues when considering powerful technologies for manipulating human thoughts and feelings.

NAVIGATING CHALLENGING CONVERSATIONS

Conversations about gendered disinformation can sometimes be uncomfortable for students and may even lead to resistance or pushback. These challenges can arise due to differing values, misunderstandings, or discomfort with discussing gender and online harm. By being proactive and prepared, educators can create a safe, supportive learning environment where students feel respected, heard, and empowered to engage.

Below are strategies for both preventing resistance and managing challenging moments when they occur:

FOSTERING EMPATHY AND COMPASSION

When engaging students in discussions about gendered disinformation, it's important to cultivate both empathy and compassion. While empathy helps students understand how others may feel when targeted by disinformation, compassion takes it a step further by encouraging them to take meaningful action. This dual approach helps build critical thinking, emotional intelligence, and a sense of responsibility in combating online harm.

Here are ways to differentiate and build on empathy and compassion in the classroom:

EMPATHY: UNDERSTANDING THE IMPACT

Empathy involves encouraging students to imagine what it's like to experience harm or harassment. This helps them emotionally connect to the issue and see the real-world consequences of gendered disinformation. Strategies for building empathy include:

- **Storytelling:** Share real or hypothetical stories of individuals targeted by gendered disinformation. Ask students, "How do you think this person might be feeling?"

- **Real-life examples:** Show age-appropriate case studies, such as instances of mis- or disinformation targeting female public figures, and discuss how this might affect their confidence, safety, and reputation.

COMPASSION AND SELF-AWARENESS: FOUNDATIONS FOR EFFECTIVE ACTION

Empathy means feeling what someone else is feeling, which can sometimes be overwhelming if their pain is intense. This could lead to feelings of paralysis - not being able to think through what you can do to help. Compassion is different — it involves caring about someone's struggle but staying calm and clear-headed, which makes it easier to think clearly and take helpful action. Compassion builds on empathy by motivating students to act against online harm. This can foster a critical and proactive mindset, where students feel empowered to create positive change. Compassion and self-awareness can work well together to enable skillful responses to provocative or difficult situations.

Strategies for encouraging compassion include:

- **Perspective taking and brainstorming solutions:** After discussing real examples of gendered disinformation, ask students: "What would I feel like if this happened to me/my friend?"; "What could you do to help stop this from spreading?" Encourage them to think about small actions, like correcting misinformation or reporting harmful content.

Strategies for encouraging self-awareness include:

- **Noticing exercises:** Taking time to pause, breathe and reflect on their thoughts, feelings, and actions. This can be helped by journaling, talking with someone they trust about reactions to provocative situations, or simply noticing how they feel in different situations.

BE PREPARED

Think in advance about what conversations could happen if gendered disinformation comes up in the classroom, and how to respond.

Before moving onto the next section, try to think about how you would approach these challenges. What other examples can you think of?

CHALLENGE	APPROACH
"Isn't this just free speech?	
"Aren't men targeted too?"	
Your example:	

STRATEGIES FOR HANDLING THESE CONVOS

As you read through these strategies, think about how they might change your approach from the examples above.

1. STOP, NOTICE, THINK, AND CAREFULLY REDIRECT

- Acknowledge their feelings and thank them for sharing.
- Avoid judgments and use empowering language.
    - Remember that your participants are members of families and communities that may have different values and beliefs than you. It's important not to make statements or suggestions that conflict significantly with family or cultural values that may come across as judgmental.
- Do not minimize or ignore the situation.

2. BE A GOOD LISTENER REDIRECT

- Maintain privacy (for example, taking students to the side rather than addressing in front of a group).
- Let youth lead and ask questions, ensuring you're avoiding judgement.
    - Helpful phrases like "What are your thoughts?", or "Why are you asking?" can help a participant express their feelings.

CHALLENGE	SUGGESTED APPROACH
"Isn't this just free speech?	Use empowering language and data: "While free speech is important, we also need to think about the difference between free speech and harmful disinformation that spreads hate or silences others." Provide examples and comparisons: Discuss the concept of harmful speech using relatable analogies, like comparing disinformation to shouting "fire" in a crowded theater.
"Aren't men targeted too?"	Recognize that disinformation affects everyone but emphasize unique patterns in gendered disinformation. Acknowledge and validate: "You're right. Disinformation can affect everyone. Let's talk about the different ways it targets various groups." Avoid minimizing or ignoring the question: Instead of dismissing the concern, steer the conversation toward exploring patterns of targeted disinformation. "All people experience disinformation, but research shows that women, especially in politics or science, face unique types of attacks that focus on credibility or appearance. Why do you think that might be?" Be a good listener: If a student feels strongly about this, listen actively and ask follow-up questions to better understand their perspective before redirecting.

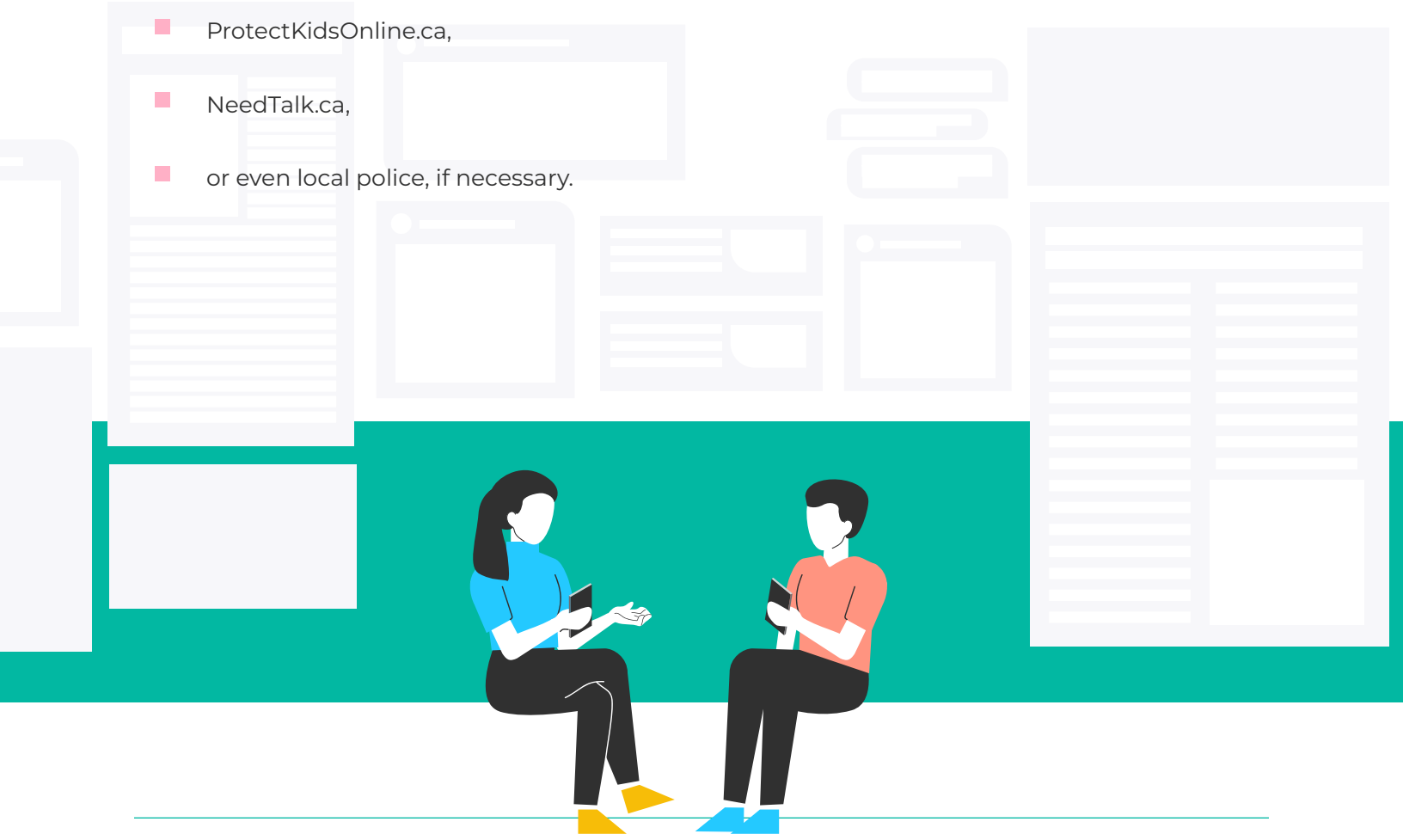
What's next?

The fight against gendered disinformation isn't just about stopping fake news—it's about creating a digital world where everyone can participate safely. By teaching media literacy, supporting those in-need, holding platforms accountable, integrating digital education into everyday learning, and fostering inclusive spaces, we can make a real difference. If we work together — across schools, families, and communities — we can build resilience against disinformation and create a more just and informed society.

It's okay to ask for help. Cyberbullying and gendered disinformation can feel overwhelming, but there are people and organizations ready to support you. By speaking up, we help make the internet a safer place for ourselves and others.

If the situation is serious, such as threats or ongoing harassment, you can also report it to:

- NeedHelpNow.ca,
- Cybertip.ca,
- the Canadian Centre for Child Protection (Protectchildren.ca),
- ProtectKidsOnline.ca,
- NeedTalk.ca,
- or even local police, if necessary.



Glossary

The digital world is constantly evolving. **Use the space below to add new terms and concepts you encounter as you deepen your understanding of this topic.**

Word	Definition
Echo chamber	A space, often online or on social media, where people only hear ideas and opinions that match their own. Because everyone shares similar views, different perspectives are rarely seen or considered. This can make someone's beliefs feel more true or more popular than they really are.
Confirmation bias	The tendency to pay more attention to information that supports what we already believe – and to ignore or dismiss anything that challenges it. It can affect how we search for, interpret, and remember information.
Misinformation	Untrue content that is spread by people who believe that it is true. Misinformation could be spread innocently, or to cause harm.
Disinformation	Untrue content that is spread by people who know that it is untrue. Disinformation is always spread knowingly and deliberately to cause harm.
Doxxing	When someone shares private information (like a home address or phone number) online to intimidate or harm a person.
Deepfake videos	Fake videos created with artificial intelligence (AI) to make it look like someone is doing or saying something they never did. These are often used to spread false, damaging information about women.
Gendered disinformation	False or misleading information designed to harm people based on their gender. It can take many forms, from online harassment, controlling behaviours, and manipulated images to false narratives that undermine the credibility of women and gender-diverse individuals.

Acknowledgements

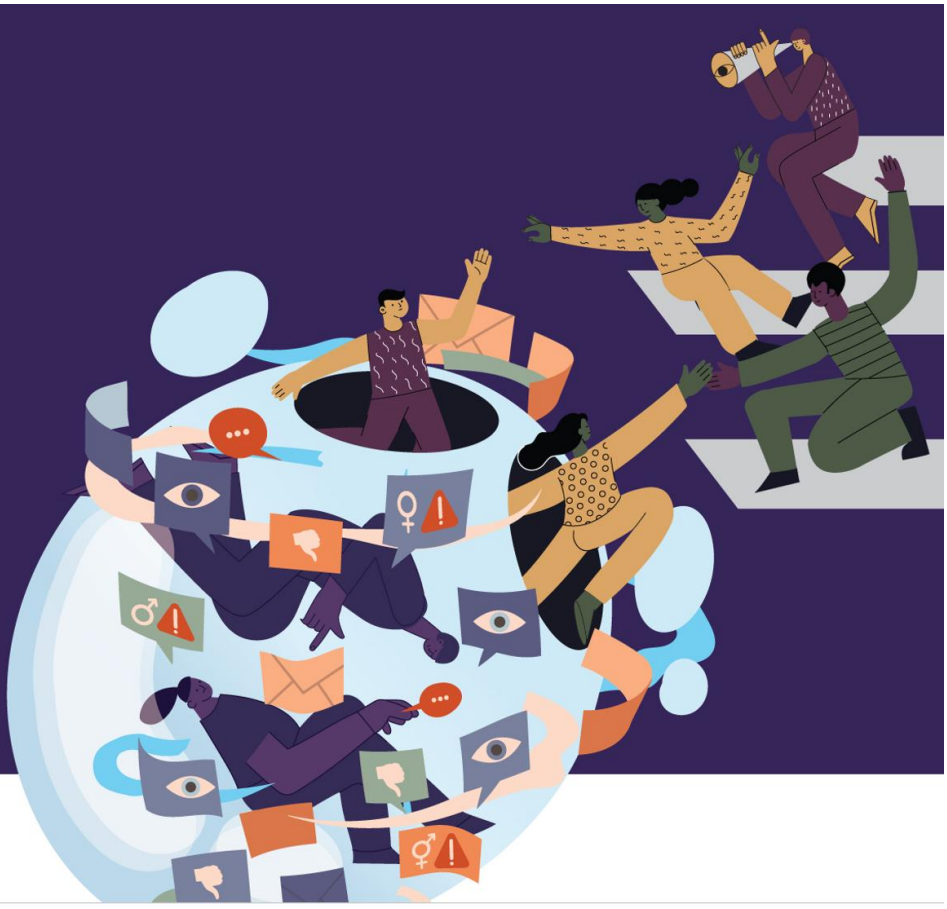
With gratitude to our advisors, reviewers, and collaborators on this work:

- Abbey Ramdeo, Actua
- Janos Botschner, Community Safety Knowledge Alliance Canada
- Janelle Fournier, Actua
- Mikayla Ellis, Actua

We also extend our sincere thanks to the members of our national community-based groups who attended special sessions related to this research and shared valuable insights to guide the development of these resources:

- National STEM Educator Community of Practice
- National Black Youth in STEM Program Youth Delegation
- National Indigenous Youth in STEM Program Youth Delegation
- Actua Network Members

Understanding and Countering Gendered Disinformation



Annex E2a

Tackling Online Gendered Disinformation: A Family Resource

CSKA Community Safety Knowledge Alliance

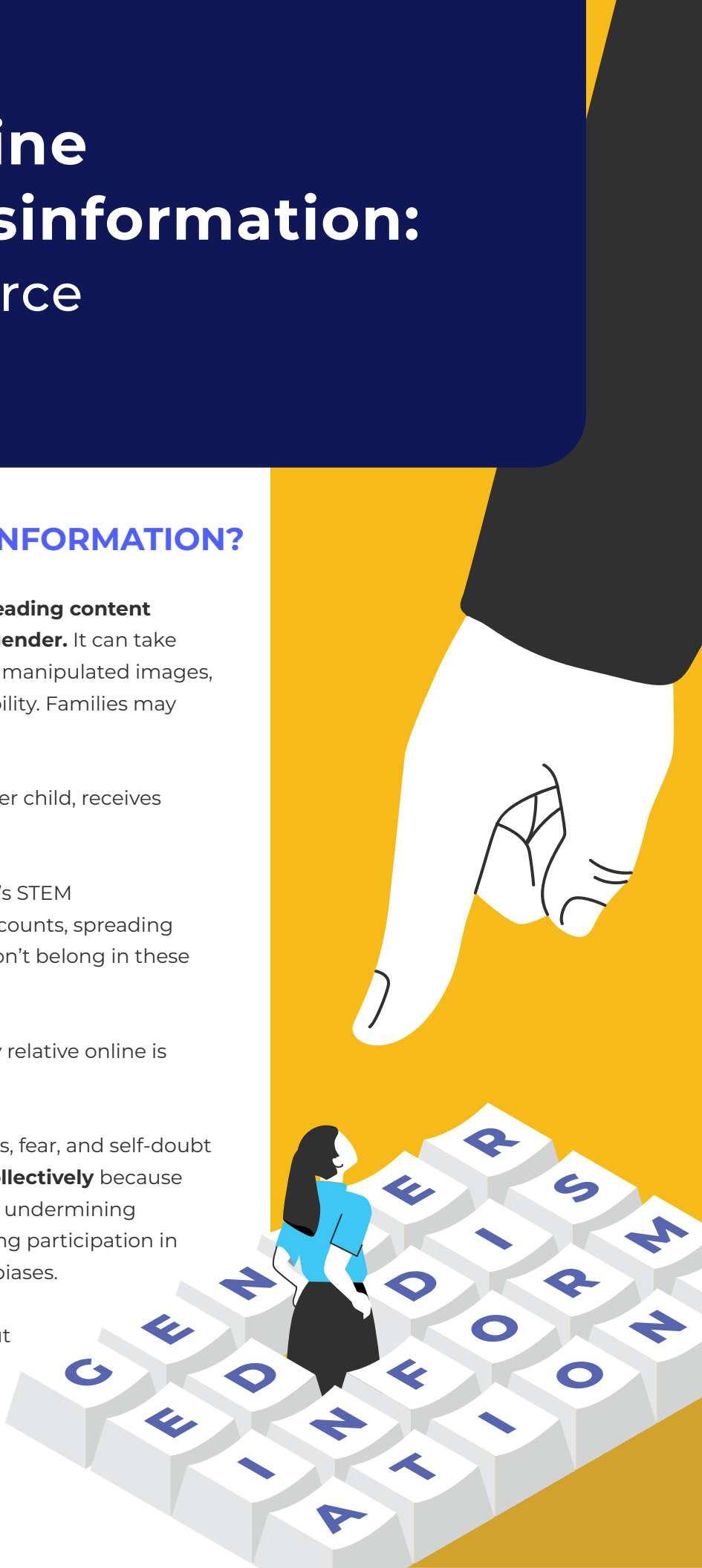
SAPPER LABS

DEVELOPED IN APRIL 2025

Tackling Online Gendered Disinformation:

A Family Resource

actua
Youth · STEM · Innovation
Jeunesse · STIM · Innovation



WHAT IS GENDERED DISINFORMATION?

Gendered disinformation is false or misleading content designed to harm individuals based on gender. It can take many forms, including online harassment, manipulated images, and false narratives that undermine credibility. Families may experience it in different ways:

- A mother supporting her transgender child, receives online hate messages and threats.
- A father posting about his daughter's STEM achievements is targeted by fake accounts, spreading false claims that girls and women don't belong in these fields.
- A sibling defending their non-binary relative online is harassed and attacked.

This type of disinformation can cause stress, fear, and self-doubt for individuals. It is also **bad for all of us collectively** because it harms society by increasing polarization, undermining confidence in credible sources, discouraging participation in public discourse, and reinforcing harmful biases.

Awareness helps. This resource talks about ways that families can recognize and respond to disinformation together - including getting help when needed.

EMPOWERING YOUR FAMILY AGAINST DISINFORMATION

1. START WITH AWARENESS

Learn the tactics:

- Doxxing (sharing private info to intimidate).
- Deepfakes & manipulated media.
- Non-consensual image sharing.
- Fake accounts, hate speech, and gendered misinformation.

Model best practices:

- Fact-checking before sharing.
- Talking openly about your own digital choices (for example, share when you block/report, why you avoid certain sites, or how you set screen time limits for yourself).

2. CREATE A SAFE SPACE FOR CONVERSATION

Safe Space:

- Validate emotions and reassure individuals it's not their fault if they're targeted.
- Let kids know they won't be punished for being honest about digital experiences.
- If you notice changes in your child's behaviour – online or otherwise – create a space to talk to them about it or consider getting help from another appropriate support network (e.g, school staff, behavioural consultants, social workers, and others).

Talk about it:

- Use media examples or digital citizenship discussions to introduce the topic.
- Encourage family members to share what they see and experience online.

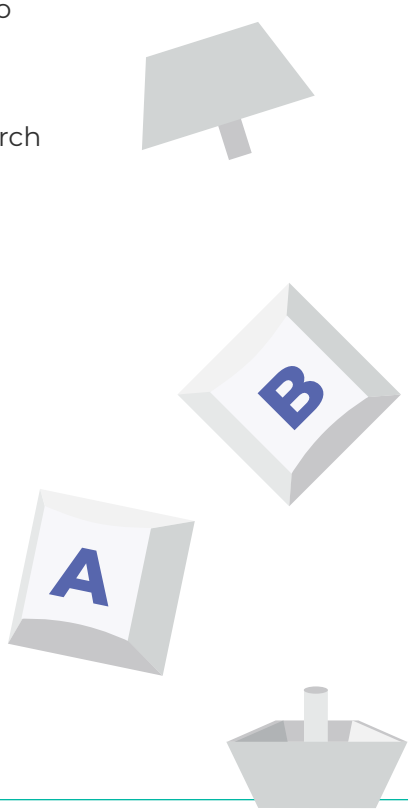
3. PRACTICE CRITICAL THINKING

Teach kids and teens to ask:

- Who created this content?
- Are they generally considered to be a credible source?
- What could their goal be?
- Is there credible evidence backing up the story - for example, is similar content being shared by other credible sources online and offline?

Help them spot:

- Posts designed to provoke strong feelings like confusion, distrust, fear, anger, or shame (e.g., fake news, manipulated images and videos):
  - **Cross-check with other sources:** Look for the same story or media on trusted news sites. If it only shows up in one place, be cautious.
  - **Check the context:** Make sure quotes, images, or videos aren't being taken out of context to mislead. Old content is often reshared to mislead—always look at when it was created.
  - **Use verification tools:** Try tools like Google Reverse Image Search to check where media originally came from.
  - **Look into the source:** Check if the site or account sharing the content is reliable and has a trustworthy history.
- Claims that reinforce harmful ideas about gender roles.
- Content that lacks or dismisses other perspectives.



4. PROMOTE SAFE ONLINE HABITS

- Use strong passwords and privacy settings.
- Block and report abusive accounts.
- Avoid sharing personal info publicly.
- Use kid-safe browser settings or tools when needed, and explore new websites as a family when possible.

5. SUPPORT EACH OTHER

- If someone is targeted, respond with care, not blame.
- Document abuse if needed and consider seeking help from relevant authorities/professionals (e.g., schools, police, legal practitioners).
- Report disinformation or harmful content together.

6. SUPPORT ONGOING LEARNING AND HEALTHY PEER RELATIONSHIPS

- Explain how critical thinking and online safety are lifelong skills to develop and practice, like fire safety, safe driving, or healthy eating.
- Like lots of skills, it helps to have people you can look to for support, and to support others along the way.
- Model and talk about the importance of friendships and other relationships that are based on respect and inclusion.
- Model and discuss the kinds of educational, volunteer and workplace interactions that reflect values of respect and inclusion.

7. AMPLIFY POSITIVE VOICES

- Follow and share accurate, empowering content.
- Model respectful online behaviour.
- Encourage children to share what they know with others.

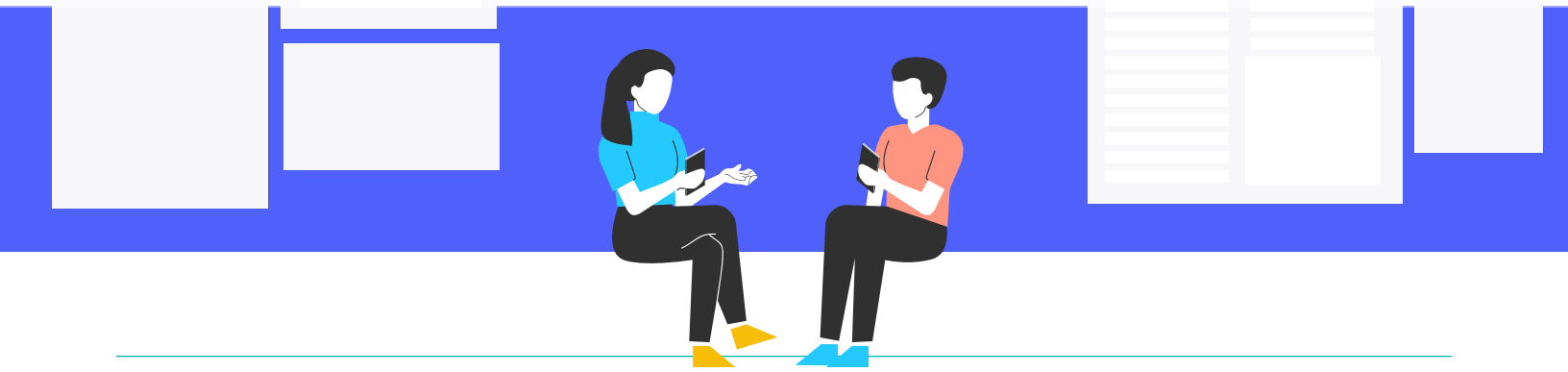
FACING DIGITAL CHALLENGES TOGETHER

Navigating online spaces can be overwhelming for young people, especially when they're facing challenges like disinformation, exclusion, or harmful content. Your words and actions can shape how your child understands and responds to these moments. Below, you can find examples of how to start supportive, open conversations – helping your child feel seen, safe, and empowered to think critically and navigate online spaces with confidence.

When you are approaching these challenges, it is important to consider the following:

- Create and model a safe, trusting space for discussion.
- Show empathy and avoid judgement.
- Encourage critical thinking and education.

Note: These scenarios can happen to any family member, but young people are more vulnerable due to their online presence and developmental stage. Whether it's a parent, grandparent, older sibling, or other caregiver, it's important for those who support young people — and each other — to be aware and ready to address these challenges.



SCENARIO	WHAT'S HAPPENING?	SUGGESTED APPROACH
AMOUNT OF TIME ONLINE	Your child has been gaming for hours, and you're noticing they seem tired or irritable.	1. **Awareness without blame:** "I noticed you've been online for a while, it's easy to lose track of time especially if you're doing something you enjoy." 2. **Empathy:** "It can be draining to be behind screens for so long. Do you want to take a break together or go for a walk?"
ECHO CHAMBERS	Your child repeats a political view they've seen shared on social media, such as claiming a specific candidate is the only 'right choice', without considering opposing viewpoints or questioning the sources of the information.	1. **Critical thinking:** "That's an interesting take. Where did you first hear about that?" 2. **Educate – encourage learning about different views:** "Sometimes we see a lot of the same kind of content online over and over again, so it can help to check out other views too. Want to explore a few different takes on this together?"
EXPOSURE TO DISINFORMATION / DEEPFAKES	Your child is upset after seeing a fake video about someone they admire or a community they belong to.	1. **Validate and empathize:** "That video seemed really intense, do you want to talk about how it made you feel?" 2. **Educate – encourage checking the facts:** "Some content is made to trick or upset us, and it can be hard to tell what's real. Let's look at some ways we can figure out what's trustworthy."
BEING TARGETED ONLINE	Your child shares that someone has been sending them mean or threatening messages, or spreading false information about them.	1. **Empathize:** "Thank you for telling me. That sounds really upsetting, and I'm here for you. You don't deserve to be treated that way." 2. **Offer Support:** "Let's go through this together and see what we can do, whether that's reporting it, saving evidence, or taking a break." 3. **Reaffirm Support:** "You are not alone."

SEEKING SUPPORT

The fight against gendered disinformation isn't just about stopping fake news — it's about creating a digital world — and a society — where everyone can participate safely. If we work together — across schools, families, and communities — we can build resilience against disinformation and create a more just and informed society.

It's okay to ask for help. Cyberbullying and gendered disinformation can feel overwhelming, but there are people and organizations ready to support you. By speaking up, we help make the internet a safer place for ourselves, our families, and others.

If the situation is serious, such as threats or ongoing harassment, you can also report it to:

- NeedHelpNow.ca
- Cybertip.ca
- the Canadian Centre for Child Protection (Protectchildren.ca)
- ProtectKidsOnline.ca
- NeedTalk.ca
- or even local police if necessary.

FAMILY GAME PLAN

- ☐ Talk about gender and online safety regularly.
- ☐ Follow trustworthy, inclusive sources together.
- ☐ Report and block harmful content – don't ignore it. Show your child how they can do the same.
- ☐ Encourage your child to come to you if something feels off.
- ☐ Normalize seeking help – from each other, from school systems or professionals and authorities.
- ☐ Learn about what your child's school might be doing to teach digital safety and to enable safe online environments. Encourage your child to participate in those positive learning environments.

Acknowledgements

This work is the product of collaboration between **Actua** and the **Community Safety Knowledge Alliance**, with **Sapper Labs Group**, and was supported, in part, through funding from Heritage Canada.

With gratitude to our advisors, reviewers, and collaborators on this work:

- Abbey Ramdeo, Actua
- Janos Botschner, Community Safety Knowledge Alliance Canada
- Janelle Fournier, Actua
- Mikayla Ellis, Actua

Actua is creating a Canada where every child has the skills and confidence they need to achieve their full potential. As a leading science, technology, engineering and mathematics (STEM) outreach organization, Actua includes over 40 universities and colleges, engaging 500,000 youth in 600 communities each year. For 25 years, Actua has focused on identifying and removing the barriers for entry into STEM and now have national programs dedicated to engaging Indigenous youth, girls and young women, Black youth, those facing economic barriers and youth in Northern and remote communities.

The Community Safety Knowledge Alliance (CSKA) is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes. Over the past decade, CSKA has conducted interdisciplinary research and engaged with change-makers on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

Sapper Labs Group (SLG) conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network. The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

CSKA

Community Safety Knowledge Alliance
Research to Practice to Alignment

SAPPER LABS

Understanding and Countering Gendered Disinformation

Annex E2b

Tackling Online Gendered Disinformation: Youth Guide

CSKA Community Safety Knowledge Alliance

SAPPER LABS

DEVELOPED IN APRIL 2025

Tackling Online Gendered Disinformation: Youth Guide

actua
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

CASE STORY: TAYLOR SWIFT & THE DANGERS OF DEEPFAKES

Content advisory: This section discusses an incident involving disturbing and sexually explicit AI-generated images. While no graphic content is shown here, the topic may be upsetting or uncomfortable for some readers. Feel free to pause or skip this section if needed.

In January 2024, sexually explicit AI-generated images of Taylor Swift began circulating across social media, particularly on the platform X (formerly Twitter). These were **deepfakes**: digitally altered images created using Artificial Intelligence to make it look like someone did or said something they never actually did. One of the images was viewed over **47 million times** before it was eventually taken down (BBC.com, 2024).

Taylor Swift is not only a major public figure in music and business – she's also recognized by many as a symbol of confidence, independence, and feminism. So when these manipulated images targeted her, the story gained wide attention. **But what happened to her is not unique**. Increasingly, this kind of content is being used to target women, girls and other gender diverse groups, often in deeply abusive and sexualized ways. Right now, the majority of deepfake content online is sexually explicit, and it overwhelmingly targets women and girls, often without their knowledge or consent (BBC.com, 2024).

While many spoke out in support of Taylor Swift, others dismissed the incident as part of the reality of being a celebrity, or minimized it by pointing out that the images were "fake." But this raises an important question: Does the fact that something is digitally created make it any less of a violation if it was shared without consent?

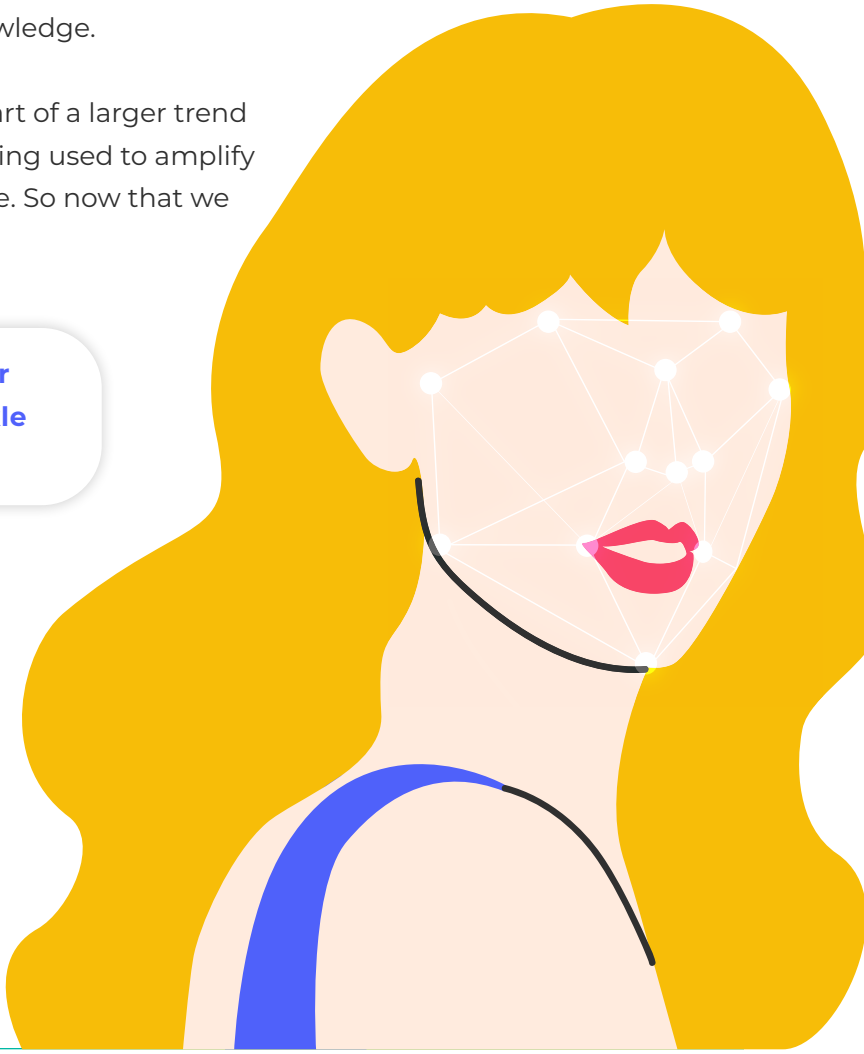
Now imagine this happening to someone you know. Or to you. Would the reaction be different?

Would it make you want to understand more about how these technologies are being misused to cause harm? Would you want to know what makes people do something like this? It turns out that this is an example of something called gendered disinformation. When false or harmful information, or fake media, about women or specific gender identities is spread online, it's an example of technology-facilitated violence.

Sadly, these kinds of incidents are far from rare. In 2023 alone, Canada's national tip line for reporting child sexual exploitation, **Cybertip.ca**, received nearly **4,000 reports of deepfake images involving youth** (Canadian Centre for Child Protection, 2024). Many of these were created and circulated without the person's knowledge.

The truth is, this isn't just about celebrities. It's part of a larger trend where emerging technologies, such as AI, are being used to amplify gender-based violence and disinformation online. So now that we know this isn't a one-off, we need to ask:

What are you going to do to empower yourself and your communities to tackle online disinformation?



WHAT IS GENDERED DISINFORMATION?

DISINFORMATION VS. MISINFORMATION

- **Misinformation** is untrue content that is spread by people who believe that it is true. Misinformation could be spread innocently, or to cause harm.
- **Disinformation** is untrue content that is spread by people who know that it is untrue. Disinformation is always spread knowingly and deliberately to cause harm.

Disinformation can be used to attack individuals or entire communities, making them feel unsafe or unwelcome in online spaces. Many people struggle to tell fact from fiction in the digital age. Algorithms push individuals toward content that reinforces their existing beliefs, creating "echo chambers" where disinformation flourishes. And, fake images and false information can spread "at the speed of cyber".

WHY DOES IT SPREAD?

THE "STICKINESS" OF DISINFORMATION

Disinformation and misinformation can be spread in a lot of ways, but social media platforms are some of the most common pathways for it. This is because social media is designed in a way to grab peoples' attention and overwhelm them with repeated stories that may or may not be true. It is so easy to repost content and help it spread further and further even if you're only sharing it to claim it's "untrue" or "silly". Some believe and share harmful content not because they want to hurt others, but because they don't realize they're being manipulated. **This is troubling because "repetition is sticky" and the more times you hear a story, the harder it is to resist believing it's true!**

Did you know? We're more likely to believe something we've seen multiple times - even if we know it's false. That's why repetition is such a strong tool for spreading disinformation.

US VS. THEM

Sometimes disinformation is used purposely to divide communities and increase polarization, pushing people into separate groups or emphasizing differences. This can make it harder to communicate well with each other and can promote conflict. When we are fighting about the truth of information, it makes it harder to get along together as a society and support one another. This is why it's important to practice communicating with people who have different opinions than your own and to think critically about the information you see online, so you can have a better chance of resisting its negative effects.

HOW DOES IT IMPACT PEOPLE?

When disinformation spreads it can make people question their abilities, limit their opportunities, or even cause them to fear speaking out, silencing their voices. Seeing negative stereotypes or false information about yourself or a community you are part of over and over can cause self-doubt, anxiety, and low self esteem. Individuals might start to question their skills or feel pressured to fit into certain expectations. For example, if you constantly hear that "women aren't good leaders," you might be less likely to put yourself forward for leadership roles at school or in your community - this hurts our capacity to function as a society. If you are repeatedly told that "people like you can't do math", you may give up on a particular career path before even starting and, as a result, society loses out on the wonderful contributions you could have made.

Gendered disinformation also affects how people treat each other, and can lead to more bullying, harassment or exclusion. This can make social media feel like an unsafe place.

TAKING ACTION

RECOGNIZE, UNDERSTAND & RESIST DISINFORMATION

Awareness of gendered disinformation, and the ability to recognize it are the first steps in countering its impact. Knowing how disinformation may appear in our daily lives ahead of time actually helps to protect us from it! **If we learn how gendered disinformation works, what tricks are used, and why people spread it, we're less likely to fall for it when we run into it online.** We'll also be in a better position to help others resist its harms, and to be able to support them if they've been affected. We can create healthier, more equal online spaces for everyone by questioning and avoiding spreading disinformation when we see it and helping others in our communities to do the same.

Below are examples of how it appears in different contexts:

- **Fake stories** – Fake news articles or social media posts that attack women, especially those in leadership roles.
- **Non-Consensual Image Sharing** – When someone shares or threatens to share private photos without permission.
- **Manipulated images & videos** – Edited pictures or deepfake videos that make it look like someone said or did something they never did.
- **Fake Accounts & Impersonation** – Creating fake profiles to spread lies, harass someone, or damage their reputation.

- **Doxxing** – When someone shares private information (like a home address or phone number) online to intimidate or harm a person.
- **Harassment, cyberbullying & hate speech** – Online attacks that try to intimidate, humiliate, or silence women and girls.
- **Memes & satire** – Jokes or cartoons that disguise harmful messages about women as "just humor".
- **Misinformation about gender roles** – Posts or comments claiming that women are naturally bad at specific tasks or in certain sectors like leadership, science, or sports.
- **Classrooms** – Comments or posts with the narrative that "girls are naturally less capable in math and science," discouraging female students from pursuing STEM education or careers.
- **Sports** – Women athletes often face public scrutiny amplified through social media, including shaming, objectification, sexist language, and debates over who is deemed eligible to compete in women's sports.

Can you think about other examples that you may have experienced or heard about?

TAKE CARE OF YOURSELF & OTHERS

To use an analogy, **it is much easier to help someone before they fall into a river than it is to pull them out when they are already downstream**. Disinformation works the same way! Building awareness of what gendered disinformation looks like and why it's harmful before it's encountered is much easier than trying to convince someone who already believes in the "fake news" that they may be wrong.

Some things you can do to empower others and contribute positively to online spaces include:

1. Understand that online spaces and digital tools can be amazing resources if used thoughtfully.

2. Recognize and resist disinformation when you see it in your own life and feed. Stop the spread with you!

3. Share what you know with others to help them be on the lookout for disinformation before it reaches them.

4. Create inclusive spaces on and offline where people can feel supported and practice communicating with others who have different lived experiences.

5. Model positive online practices like checking sources and thinking critically about information for your friends, family and communities.

6. Diversify who you follow so you aren't trapped in an "echo chamber" where you only see ideas or opinions that already match your own.

7. Support those who may be targeted by gendered disinformation by helping them connect with trusted networks and resources like family members, teachers, youth workers, or online services.

MY DIGITAL ACTIONS CHECKLIST

- ☐ I pause before sharing posts, especially if they seem emotional or extreme.
- ☐ I check where the content came from and whether it seems trustworthy.
- ☐ I report harmful or sexist content when I feel safe to do so.
- ☐ I talk to friends about what we see online, not just what's funny or viral, but what feels off or untrue.
- ☐ I take breaks from social media when it starts to feel overwhelming.

IT'S OK TO STEP AWAY!

Conversations about gendered disinformation can sometimes be uncomfortable and may have you feeling a range of emotions. **A lot of disinformation – on social media especially – is designed to cause fear, anger or confusion so it can gain traction.** It's important to recognize when it's having a negative impact on your own health and wellbeing so you can take a step back or seek support. It can be exhausting to constantly be educating others or handling harmful content in your online spaces. Action against disinformation is a team effort and that means it's not your job to fight every battle!

Cyberbullying and gendered disinformation can feel overwhelming, but there are people and organizations ready to support you. By speaking up, we help make the internet a safer place for ourselves and others.

If the situation is serious, such as threats or ongoing harassment, you can also report it to:

- NeedHelpNow.ca,
- Cybertip.ca,
- the Canadian Centre for Child Protection (Protectchildren.ca),
- ProtectKidsOnline.ca,
- NeedTalk.ca,
- or even local police if necessary.

REFLECTING ON YOUR ONLINE WORLD

Now that you've reached the end of this guide, reflect on the following:

One thing I learned that surprised me: _____

_____

One way I can create safer online spaces: _____

_____

One person I want to talk to about this: _____

_____

Glossary

The digital world is constantly evolving. **Use the space below to add new terms and concepts you encounter as you deepen your understanding of this topic.**

Word	Definition
Echo chamber	A space, often online or on social media, where people only hear ideas and opinions that match their own. Because everyone shares similar views, different perspectives are rarely seen or considered. This can make someone's beliefs feel more true or more popular than they really are.
Misinformation	Untrue content that is spread by people who believe that it is true. Misinformation could be spread innocently, or to cause harm.
Disinformation	Untrue content that is spread by people who know that it is untrue. Disinformation is always spread knowingly and deliberately to cause harm.
Doxxing	When someone shares private information (like a home address or phone number) online to intimidate or harm a person.
Deepfake videos	Fake videos created with artificial intelligence (AI) to make it look like someone is doing or saying something they never did. These are often used to spread false, damaging information about women.
Gendered disinformation	False or misleading information designed to harm people based on their gender. It can take many forms, from online harassment, controlling behaviours, and manipulated images to false narratives that undermine the credibility of women and gender-diverse individuals.

Acknowledgements

This work is the product of collaboration between **Actua** and the **Community Safety Knowledge Alliance**, with **Sapper Labs Group**, and was supported, in part, through funding from Heritage Canada.

With gratitude to our advisors, reviewers, and collaborators on this work:

- Abbey Ramdeo, Actua
- Janos Botschner, Community Safety Knowledge Alliance Canada
- Janelle Fournier, Actua
- Mikayla Ellis, Actua

We also extend our sincere thanks to the members of our national community-based groups who attended special sessions related to this research and shared valuable insights to guide the development of these resources:

- National STEM Educator Community of Practice
- National Black Youth in STEM Program Youth Delegation
- National Indigenous Youth in STEM Program Youth Delegation
- Actua Network Members

Acknowledgements

Actua is creating a Canada where every child has the skills and confidence they need to achieve their full potential. As a leading science, technology, engineering and mathematics (STEM) outreach organization, Actua includes over 40 universities and colleges, engaging 500,000 youth in 600 communities each year. For 25 years, Actua has focused on identifying and removing the barriers for entry into STEM and now have national programs dedicated to engaging Indigenous youth, girls and young women, Black youth, those facing economic barriers and youth in Northern and remote communities.

The Community Safety Knowledge Alliance (CSKA) is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes. Over the past decade, CSKA has conducted interdisciplinary research and engaged with change-makers on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

Sapper Labs Group (SLG) conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network. The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

CSKA
Community Safety Knowledge Alliance
Research to Practice to Alignment

SAPPER LABS

Understanding and Countering Gendered Disinformation

Annex E2c

Tackling Online Gendered Disinformation: Additional Resources for Educators, Families & Youth

CSKA Community Safety Knowledge Alliance

SAPPER LABS

DEVELOPED IN APRIL 2025

Tackling Online Gendered Disinformation:

Additional Resources for Educators, Families and Youth

actua
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

BEST PRACTICES

- **Safe surfing** - Practice safe surfing from malware and toxic content. Where are you getting your news?
  - Less trustworthy sites are Tik Tok, Gab, Rumble, Telegram, Zello, Truth Social, VK.
  - Reputable sites include those like CBC, Government of Canada's Science and Innovation, La Presse (in French), Radio Canada (in French), Scientific America, Nature, Science Magazine, or BBC.
- **Hot topics** - Evaluate content that covers politics, religion, crime, social issues, celebrities or conspiracies. These topic areas are often in the news and can be sensationalized or designed in ways to invoke specific reactions.
- **Feelings** - Acknowledge if information sounds too good or bad to be true. Is it controversial, negative, sensational, exaggerated and provocative? Does it invoke an emotional response? It is even more important to be skeptical in these cases.
- **Agendas** - Consider why someone posted content. Who are they trying to hurt? Are they trying to make you the victim of harm or the "tool" to harm others?

- **Profit** - Recognize that bad news sells. Is the headline and story click bait to sell ad revenue or ratings?

- **Street cred** - Check credentials of posters and websites.

- **Real source** - Assess primary source reliability and credibility. Even a highly referenced article can fall apart like a house of cards when the root source material is fraudulent.

- **Truth** - Verify if information is supported by the majority of experts and authoritative sites.

- **Peer review** - Search the claim to see whether it has been fact-checked or debunked already. Check the date.

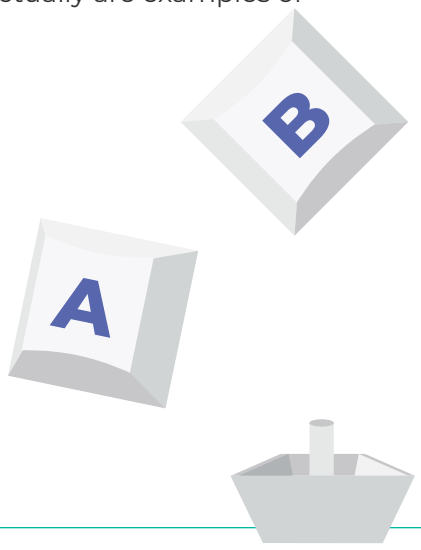
- **Deepfakes** - Utilize tools like Google reverse image search to help discover the source of an image and its possible variations.

- **URLs** - Investigate for domain manipulation, which is widespread.

    - For example, what looks like an .mil domain, followed by .co or “lo” is likely a fake or deceptive site. If you are noticing a slightly variant version of a well-known URL, do some investigating. Does the SSL certificate (Secure Sockets Layer certificate) check out? An SSL certificate is a digital certificate that authenticates a website’s identity and enables a secure, encrypted connection between a web server and a user’s browser (An example of a free SSL checker: https://sslcertificatechecker.com/).

- **Signal manipulation** - Stop to consider sites that use words in their titles designed to makeyou inherently trust them, such as democratic, research, peace, or truth.

    - E.g., Democratic People’s Republic of Korea [DPRK], Global Research Canada, Internet Research Agency, World Peace Council, Foundation to Battle Injustice, Canadian Peace Congress are all sites named in ways to signal reliability but actually are examples of untrustworthy or biased organizations.



PROTECT YOURSELF AND TAKE ACTION

- **Recognize the trick** – Learn the common tactics used in gendered disinformation, like fake statistics, emotional manipulation, and edited images. When you know the tricks, they lose their power.

- **Think before you believe and share** – Don't be a spreader! When you see a shocking claim, pause. Ask yourself: Who is saying this? What's their motive? Where's the evidence? How do I contribute to the spread of disinformation by sharing this with others?

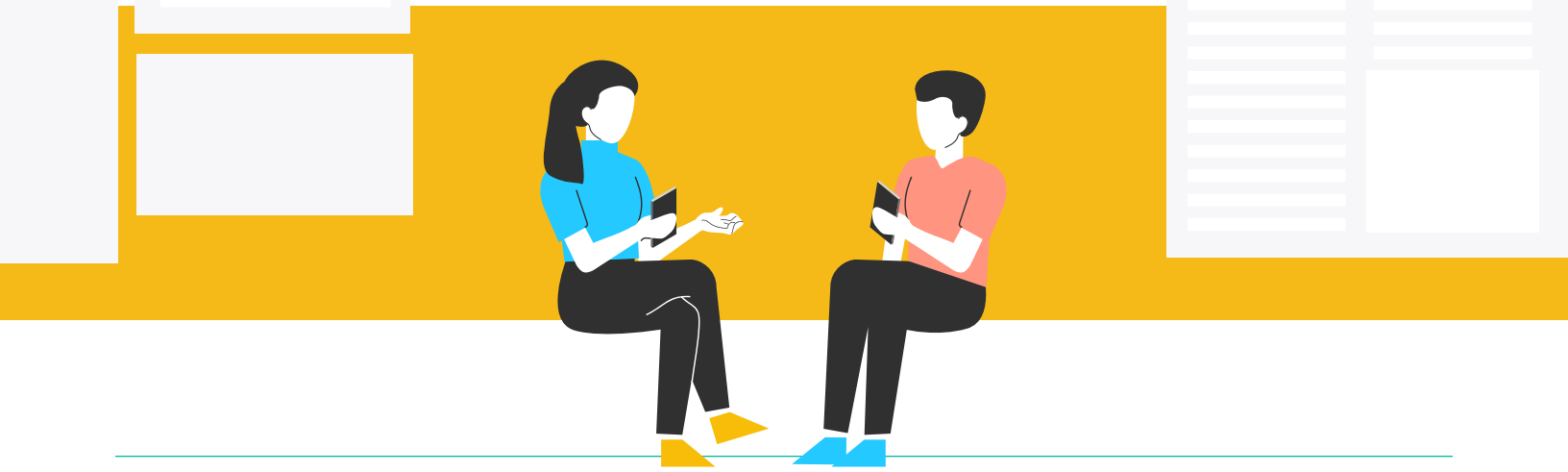
- **Check your own reactions** – Disinformation is designed to make people feel angry, scared, or doubtful. If a post sparks a strong emotion, take a step back and analyze it before reacting.

- **Expose the tactics** – Talk about gendered disinformation with friends and family. If you call out false claims when you see them, you help others develop mental immunity, too!

- **Surround yourself with positive and reliable information** – Follow pages and influencers that share accurate, empowering, and fact-based information.

- **Report harmful content** – to parents, teachers, counsellors or other trusted adults. Social media platforms also have tools for reporting harassment or false information.

- **Friends can also support each other** – You can do this by creating a trusted space to talk about the challenges and benefits of using social media and other online platforms.



EXAMPLES OF SUPPORTS AND TOOLS

Below is a non-exhaustive list of tools currently available. They illustrate ways in which technology can have a positive role to play in tackling online gendered disinformation.

Available tools and resources may change over time and their efficacy will evolve with changing online landscapes. This does not serve as an endorsement of any of the tools below, it is a list of current examples.

TOOLS FOR ANALYTICS/AUTOMATED DETECTION

WEVERIFY | **DUCKDUCKGOOSE** | **DEEPFAKEPROOF**

WHAT THEY DO	- Content verification, tracking, and debunking (WeVerify); AI powered deepfake detection for images/videos/audio (DuckDuckGoose); - Helps users identify deepfakes while browsing the web (DeepfakeProof).
HOW TO USE THEM	- Chrome Plugin (WeVerify); - Upload files via a browser to DuckDuck Goose; - A real-time deepfake detection plugin for Chrome (DeepfakeProof).
SUBSCRIPTION	- Free/Open source platform (WeVerify); - Subscription Required (DuckDuckGoose); - Free Chrome Plug-in (DeepfakeProof).
EXAMPLE	A fake nude image of a female politician is detected and debunked.

PLATFORM/CONTENT GENERATOR TOOLS

SYNTHID (DIGITAL WATERMARKING)

WHAT IT DOES	Watermarks and identifies AI generated content by embedding digital watermarks directly into AI generated images, audio, text, or video.
HOW TO USE IT	Integrated into AI-generated media, detected by compatible tools.
SUBSCRIPTION	Available via Google Cloud's AI tools (Google DeepMind).
EXAMPLE	A fake image of a female CEO is debunked using SynthID detection.

TRANSPARENCY

GENDER-SENSITIVE MONITORING

WHAT IT CAN DO	One can utilize AI tools (e.g., Reality Defender), social network analysis (Hoaxy, MeVer) and qualitative methods to track gendered disinformation.
HOW IT CAN BE USED	Quantitative / qualitative research to identify gendered attacks online.
SUBSCRIPTION	Varies — some tools are free, others require paid access.
EXAMPLE	Through gender-sensitive monitoring, a researcher is able to show that women candidates face twice as many disinformation attacks as men.

ENHANCED USER REPORTING FOR HARMFUL CONTENT

WHAT IT CAN DO	Improve response time and categorization of gendered disinformation reports.
HOW IT CAN BE USED	As a mass-reporting campaign.
SUBSCRIPTION	Unknown, would depend on platform implementation.
EXAMPLE	A journalist targeted by deepfakes reports it to an enhanced moderation system.

Acknowledgements

This work is the product of collaboration between **Actua** and the **Community Safety Knowledge Alliance**, with **Sapper Labs Group**, and was supported, in part, through funding from Heritage Canada.

With gratitude to our advisors, reviewers, and collaborators on this work:

- Abbey Ramdeo, Actua
- Janos Botschner, Community Safety Knowledge Alliance Canada
- Janelle Fournier, Actua
- Mikayla Ellis, Actua

Actua is creating a Canada where every child has the skills and confidence they need to achieve their full potential. As a leading science, technology, engineering and mathematics (STEM) outreach organization, Actua includes over 40 universities and colleges, engaging 500,000 youth in 600 communities each year. For 25 years, Actua has focused on identifying and removing the barriers for entry into STEM and now have national programs dedicated to engaging Indigenous youth, girls and young women, Black youth, those facing economic barriers and youth in Northern and remote communities.

The Community Safety Knowledge Alliance (CSKA) is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes. Over the past decade, CSKA has conducted interdisciplinary research and engaged with change-makers on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

Sapper Labs Group (SLG) conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network. The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

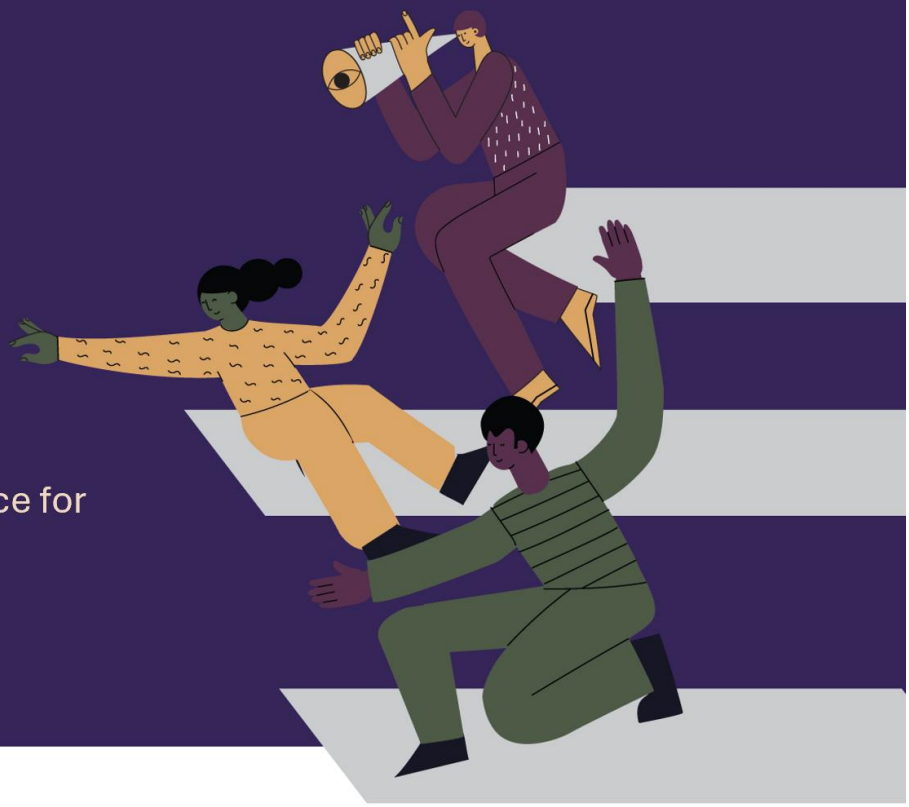
actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

CSKA

Community Safety Knowledge Alliance
Research to Practice to Alignment

SAPPER LABS

Understanding and Countering Gendered Disinformation



Annex E3

Gendered Disinformation: A Resource for Police & Human Service Agencies

CSKA Community Safety Knowledge Alliance

SAPPER LABS

Understanding Gendered Disinformation

Online abuse and fake information targeting gender

A Short Guide for Police and Human Service Agencies

CSKA Community Safety Knowledge Alliance
Research to Practice to Alignment

SAPPER LABS

This guide provides police and human service agencies with information to help understand and collaborate on countering gender-based disinformation and abuse. It is intended as an introduction – a starting point – for shared learning and collaboration.

A set of companion resources for parents and youth, and for educators, have been developed in collaboration with Actua (www.actua.ca).

The source material for this resource, with detailed information on gendered disinformation, can be found in the following report:

> **Botschner, J., Cioffi, G., McMahon, D., Ollinger, J., Sylvestre, B., Kotak, R. & Corley, C. (2025). Mobilizing awareness and building resilience to counter gendered disinformation. Ottawa ON: Community Safety Knowledge Alliance.**

This work is the product of a collaboration between:

CSKA Community Safety Knowledge Alliance
Research to Practice to Alignment

SAPPER LABS

With funding from:

Canada

Digital Citizenship Contribution Program, Canadian Heritage

Acknowledgements:

We thank the following individuals who provided feedback on an earlier version of this guide:

Delta Police Department	Greater Sudbury Police Service	Sudbury YWCA
André Cruz *Communications Assoc.*	Det. Sgt. Adam Demers *Criminal Investigations/Intimate Partner Violence*	Marlene Gorman *Executive Director*
Cst. Derek Defrane *Domestic Violence Unit*	Dan Gelinas *Community Mobilization Liaison*	
Kim Gramlich *Mgr., Victim Services*	Det. Sgt. Lee Rinaldi *Major Sex Crimes*	
Sgt. Alex Quezada *Vulnerable Sector Unit*		

Online Gendered Abuse and Disinformation

Understanding the Problem

What "Gendered Disinformation" Means

Gendered disinformation involves the misuse of digital platforms and technologies to harm individuals or groups of women, girls, or gender non-conforming persons, although it can also occur through other means[1].

It involves spreading false or misleading information, drawing on sexist, misogynistic, and transphobic stereotypes, often portraying targets are untrustworthy, immoral, or not deserving of credibility. Content may include explicit threats of physical harm to targets or their loved ones, or hints of potential harassment or harm. This can also be done by indirectly inciting others to participate in abuse against the targeted individual(s).

This is a widespread global problem. A recent study found that, globally, 85 percent of women had witnessed or experienced online gender-based violence and 38 percent had been personally impacted by it[2]. The issue is probably more widespread than this.

Gender based violence against Indigenous women and girls is a serious problem in Canada – they face greater levels of intimate and non-intimate partner violence than non-Indigenous females[3].

A recent study by the Canadian Women's Foundation found that thirty percent of Indigenous women encounter unwelcome behavior online[4]. The same study found that 20 percent of Canadian women experience some form of online harassment. Individuals who have identities that cut across more than one gender/sexual orientation and/or racialized category are likely at higher risk of being exposed to or targeted by gendered disinformation.

Gendered Disinformation Causes Harm

The psychological, safety and reputational impacts on victims can be significant. Many experience ongoing anxiety, social withdrawal, loss of professional opportunities, or direct threats to their security. In 2015, the Criminal Code of Canada was amended (at section 162.1) making it an indictable offence, punishable by up to 5 years in prison, for sharing, selling, making available or advertising sexually explicit images of someone without their consent. Many believe that there is

---

[1] Gendered disinformation can be spread outside of digital environments. While its victims are largely female-identifying and gender non-conforming persons, male-identifying persons may also (less frequently) be the targets. Gendered disinformation is not confined to intimate partner violence – it can be a form of violence against women and gender diverse persons more generally.

[2] Economist Intelligence Unit (2020) data from 2019-2020, reported by Jankowicz, et al. (2024)

[3] Between 2019 and 2020, based on 2024 SDG Gender Index, published by Equal Measures 2030 (a global coalition of NGOs that use data and evidence to address gender equality)

[4] Canadian Women's Foundation (n.d.)

more work to be done to improve the prevention and response to gendered disinformation and other forms of online abuse.

Gendered disinformation often targets individuals.

This form of disinformation can be part of taking revenge against, or controlling, someone. It can also be used to interfere with advocacy, silence different views, and discourage participation in civic life. It often goes beyond one-off insults or slurs, to coordinated campaigns using fake accounts, bots, and manipulated media (such as "deepfaked" content) to expand harm.

While public figures (e.g.,politicians, celebrities, etc.) are frequent targets, anyone can be targeted. For example, it is common for gendered disinformation to appear in schools as bullying and harassment.

Larger social impacts include decreased civic participation of women and gender-diverse people, normalizing gender-based abuse, and harming our ability to communicate.

Gendered disinformation can be used to attack groups in order to disrupt Canadian society.

Gendered disinformation can also be a tool of foreign interference. Foreign governments, criminal organizations or ideological groups may use attacks on women, gender diverse issues or gender equity to sow polarization and undermine trust in our democracy and institutions.

Why This Matters for Police and Community Agencies

This isn't just a technology issue – it's a community safety and wellbeing issue.

Perpetrators of intimate partner violence and coercive control can spread false information, fake images as part of their abuse of victims. Awareness of these tactics may help police, justice and human services better understand the risks and respond appropriately.

Understanding that gendered disinformation can be used as part of efforts to undermine our democracy can help police and others flag potential national security risks.

Responding effectively requires awareness, coordinated support, and trauma-informed practices.

Common Terms and Techniques

Examples of disinformation and tech-based abuse include:

- **Fake stories** – Fake news articles or social media posts that attack individuals, such as former partners/spouses, or those in public or leadership roles.
- **Non-consensual image sharing** – Can include posting or re-posting intimate images that were meant to be private or exclusive to a partner. It can also involve uploading sexual photos or videos of an ex-partner to social media or pornographic websites without their consent.
- **Manipulated images & videos** – Edited pictures or **"deepfake"** videos that make it look like someone said or did something they never did. Commonly encountered situations include non-consensual, out-of-context, sharing of manipulated or real photos/fake explicit content.
- **Misinformation about gender roles** – Posts or comments claiming that women are naturally bad at leadership, science, or sports.
- **Harassment & cyberbullying** – Online attacks that try to intimidate, humiliate, or silence.
- **Fake accounts & impersonation** – Creating fake online profiles to spread lies, harass someone, or damage their reputation. When this involves creating many fake accounts, or taking over

existing accounts to make it look like other people – individual influencers or crowds – agree with a fake story, it is called **astroturfing**.

- **Memes & satire** – These are jokes or cartoons that disguise harmful messages about their targets as "just humor."
- **Doxxing** – Broadly sharing private information (like a home address or phone number) online to intimidate or harm a person. Sometimes, this can lead to offline intimidation or violence.
- **Surveillance and manipulation of "smart" technology** – For example, using commercially available tracking devices, to monitor someone's movement; or manipulating home or vehicle systems to intimidate someone.

Effective responses may include:

- **Validating** the victim's experience and avoiding minimization.
- **Documenting** online content for evidence.
- **Coordinating** with digital forensic or cybercrime units.
- **Referring** victims to community-based supports.
- **Taking steps to protect** victims from further digital exposure (e.g., safety planning, privacy settings, reporting abuse to platforms).
- **Collaborating** across the service system to ensure that everyone who has been involved is connected to the right source of support and/or accountability.
- **Promoting community awareness** supports prevention, accountability and resilience.

Building Effective Police and Community Partnerships

Police, human service and community-based organizations can work together to better support victims and reduce harm. Agencies can:

- **Share knowledge** about digital abuse trends – internally and to improve public awareness;
- **Encourage and develop appropriate reporting pathways** – for example, around potential criminal code violations such as the sharing of nude images of minors;
- **Establish referral pathways** and warm handoffs;
- **Provide joint training** on gendered disinformation, trauma-informed and culturally safe practices; and
- **Engage with prosecution service and policy makers** to improve protections against, and responses to, gendered disinformation.

Appendix: Additional Information and Resources[5]

Examples of tech-facilitated abuse and gendered disinformation in the news

Threats, harassment and online hate driving women out of policitcs, MPs warn

By Jesmeen Gill - The Canadian Press
March 8, 2025
Source: https://globalnews.ca/news/11073007/threats-harassment-and-online-hate-driving-women-out-of-politics-mps-warn/
Excerpt: As longtime Liberal MP Pam Damoff prepares to leave politics when the next federal election is called, she is wistful but open about what is driving her to leave a career she has had for more than a decade. Vocal about the misogyny and threats she faced during her time in government, she wants public safety officials to take these threats more seriously. "We've seen a shift in how people treat politicians, and I really worry that at some point, someone will be injured or killed," Damoff said in an interview.

Across globe, women battle 'gendered disinformation'

AFP News
March 23, 2023
Source: https://www.news.com.au/breaking-news/across-globe-women-battle-gendered-disinformation/news-story/26f0a9c36d1b9bc421579cef9ffb3802
Excerpt: Fake photos showing Ukraine's first lady sunbathing topless, incorrect video subtitles defaming Pakistani feminists for "blasphemy", slow-motion clips falsely depicting "drunk" female politicians -- a barrage of disinformation targets women in the public eye.... Last year, a fake image of Ukraine's First Lady Olena Zelenska lying topless on a beach in Israel was shared widely on Facebook, triggering criticism that she was having fun while her war-torn country was suffering. A reverse image search by AFP showed the woman in the photo was, in fact, a Russian television presenter. Former American first lady Michelle Obama and current French first lady Brigitte Macron have also been targeted in false online posts that claimed they were born as men. The disinformation sparked an avalanche of mockery and transphobic remarks. New Zealand's Jacinda Ardern, who announced her resignation as prime minister in January, is another prominent figure that faced a torrent of disinformation about her sex.... When Germany's current foreign minister, Annalena Baerbock, was running for chancellor in 2021, she was the subject of frequent disinformation campaigns which raised questions about whether she was fit for the job. One of them featured images of a nude model purporting to be of her, alongside suggestions that she had engaged in sex work.

[5] The resources listed here are examples of currently available technologies. These tools and services are meant to illustrate options for further exploration; their listing here is not meant as an endorsement of any particular resource. In addition, resources may change and new tools are becoming available all the time.

'Stalked within your own home': Woman says abusive ex used smart home technology against her

In our hyperconnected world, 'tech abuse' is becoming a growing problem

Makda Ghebreslassie - CBC News

November 1, 2018

Source: https://www.cbc.ca/news/science/tech-abuse-domestic-abuse-technology-marketplace-1.4864443

Excerpt: While smart technology — web-controlled devices like locks, lights, thermostats and cameras — can provide convenience and a sense of security for some, these tools are increasingly being used by others to monitor, harass, stalk and intimidate. "He was able to monitor me, you know, using the security surveillance cameras, even remotely, from thousands of miles away," she said. "You're never outside the reach of your abuser." During an especially strained time in their relationship, while they were living apart, Nijem said [her ex-partner] maintained control of the house and would use it to terrorize her. "In the middle of the night, I'm awoken, and my dogs are awoken, by this blaring music over the audio system. You have lights flickering on and off, TVs going on and off," she said. "It's almost as if the house is haunted," Nijem said. "It is only done to cause you trauma, to cause fear, to cause anxiety."

Explicit fake images of Taylor Swift prove laws haven't kept pace with tech, experts say

Experts say laws must target developers, social media companies and individual users

Rhianna Schmunk – CBC News

January 26, 2024

Source: https://www.cbc.ca/news/canada/taylor-swift-ai-images-highlight-need-for-better-legislation-1.7096094

Excerpt: Explicit AI-generated photos of one of the world's most famous artists spread rapidly across social media this week, highlighting once again what experts describe as an urgent need to crack down on technology and platforms that make it possible for harmful images to be shared. Fake photos of Taylor Swift that depicted the singer-songwriter in sexually suggestive positions were viewed tens of millions of times on X, previously known as Twitter, before being removed. One photo, shared by a single user, was seen more than 45 million times before the account was suspended. But by then, the widely-shared photo had been immortalized elsewhere on the internet.

Police arrest 12 people after investigation into 'non-consensual sharing of intimate content'

Thunder Bay police say 117 alleged victims in Thunder Bay, Canada and abroad

Michelle Allan – CBC News

February 5, 2025

Source: https://www.cbc.ca/news/canada/thunder-bay/police-arrest-12-people-after-investigation-into-non-consensual-sharing-of-intimate-content-1.7451252

Excerpt: "Multiple women in Thunder Bay had been having their images shared through groups in an online chat platform," said police in a release. After a lengthy cyber crime investigation, police identified a total of 117 alleged victims in Canada and in at least three other countries, said TBPS. The twelve people accused face a total of 172 charges, said police. The majority of the charges laid were Distribution of Intimate Images Without Consent.

Reports of sharing intimate photos without consent increased in Regina and Saskatoon, police say

Saskatoon Police Service received 28 reports in 2021

Theresa Kliem – CBC News

March 1, 2022

Source: https://www.cbc.ca/news/canada/saskatchewan/distribution-of-non-consensual-intimate-images-regina-1.6367480

Excerpt: Police in Regina and Saskatoon say there were more complaints about intimate images being shared without consent in 2021 than in previous years. Sixteen people made these reports to the Regina Police Service (RPS) last year compared to 11 in 2020 and eight in 2019, according to the force's 2021 crime statistics.

Seeing is believing the real and present danger of fake AI images

In December, the Winnipeg Police Service were investigating reports of AI-generated nude photos of underage students circulating at Collège Béliveau, a Grade 9-12 high school in Windsor Park

Jen Zoratti – Winnipeg Free Press

February 10, 2024

Source: https://www.winnipegfreepress.com/arts-and-life/2024/02/10/seeing-is-believing-the-real-and-present-danger-of-fake-ai-images

Excerpt: The speed and ease with which these images can be created and spread is also alarming; one doesn't even need to have a mastery of Photoshop anymore.... And yet, despite this rapid acceleration in technology, it seems as if we're still stuck in 2014 when it comes to the law. ... Per a recent Canadian Press story about Canadian provinces playing catchup on this file, Manitoba is one of eight provinces that do indeed have intimate image laws, but ours don't refer to altered images. That needs to change, and fast. We cannot afford to have the creation and distribution of sexually explicit AI-generated images dealt with the same way online sexual harassment has traditionally been dealt with, which is to just tell women to "stay off the internet."

An interview between CBC Newsworld host Arti Pole and Ritesh Kotak on this incident and topic may be found at:
https://cbchls.akamaized.net/delivery/news/2023/12/18/ritesh-kotak-dec18-17-01-13/ritesh-kotak-dec18_5000kbps.mp4

Sources of Information and Support

- **IPV & Digital Abuse:**
    - Tech Safety Canada (https://www.techsafetycanada.ca/);
    - BC Society of Transition Houses (https://bcsth.ca/);
    - Justice Canada (https://www.canada.ca/en/women-gender-equality/gender-equality/gender-results-framework/gender-based-violence-access-justice.html);
    - Canadian Women's Foundation (https://canadianwomen.org/);
    - National Network to End Domestic Violence (US) (https://nnedv.org/)
    - Government of Canada: (https://www.canada.ca/en/security-intelligence-service/corporate/publications/the-evolution-of-disinformation-a-deepfake-future/developing-grounded-human-rights-centered-responses-to-deepfakes-synthetic-media-and-audiovisual-generative-ai.html)
- **Hate & Extremism Monitoring:**
    - Canadian Centre for Cyber Security (https://www.cyber.gc.ca/en/);
    - Canadian Anti-Hate Network (https://www.antihate.ca/about/).
- **Foreign Interference & Transnational Repression:**
    - Canadian Security Intelligence Service (CSIS) (https://www.canada.ca/en/security-intelligence-service/corporate/publications/foreign-interference-and-you/foreign-interference-and-you.html);
    - Government of Canada, Department of Democratic Institutions (https://www.canada.ca/en/democratic-institutions/news/2025/01/protecting-canadas-democratic-institutions-and-processes-from-foreign-interference.html/)
    - Citizen Lab (https://citizenlab.ca/2024/12/the-weaponization-of-gender-for-the-purposes-of-digital-transnational-repression/).

Sample Technology Options for Human Service Organizations and Others

ANALYTICS/AUTOMATED DETECTION

Note: Weblinks have not been provided as these may change over time. Subscription fees, where indicated, are current, as of April, 2025. New solutions are regularly being developed and tested.

1. **Sentinel Deepfake Detection System**
    - **What it does**: AI detection platform that works with governments, media, & defence agencies to protect democracies from disinformation campaigns, synthetic media & information operations.
    - **How to use it**: Users can report gendered deepfakes for review.
    - **Subscription**: No public access; used by governments, media, & defence agencies.
    - **Example**: A deepfake targeting a female journalist is flagged & removed before going viral.

2. **WeVerify, DuckDuckGoose, DeepfakeProof**
   - **What they do**: Content verification, tracking, & debunking (WeVerify); AI powered deepfake detection for images, videos, & audio (DuckDuckGoose); Helps users identify deepfakes while browsing the web (DeepfakeProof).
   - **How to use them**: Chrome Plugin (WeVerify); Upload files via a regular browser to DuckDuck Goose; As a real-time deepfake detection plugin for Chrome (DeepfakeProof).
   - **Subscription**: Free/Open-source platform (WeVerify); Subscription Required (DuckDuckGoose); Free Chrome Plug-in (DeepfakeProof).
   - **Example**: A fake nude image of a female politician is detected & debunked.

3. **Reality Defender**
   - **What it does**: Equips enterprises, governments, & platforms with the tools to detect AI generated or manipulated content in real time.
   - **How to use it**: Upload content to the software for real-time video identity, image & text authentication.
   - **Subscription**: Subscription required.
   - **Example**: A fake video targeting a women's rights activist is debunked before being used in a smear campaign.

4. **MeVer: Verification, Media Analysis, & Retrieval**
   - **What it does**: Developing technologies & services for understanding, searching, & verifying media content
   - **How to use it**: Journalists & researchers analyze disinformation content & networks.
   - **Subscription**: Offers resources (tools, software, & datasets) via GitHub & other repositories.
   - **Example**: A smear campaign against female journalists is traced to coordinated disinformation actors.

5. **RAND's Countering Truth Decay Initiative**
   - **What it does**: RAND researchers are studying the causes, consequences, & means of countering truth decay.
   - **How to use it**: Free resource.
   - **Subscription**: Research available on RAND's website for free.
   - **Example**: A journalist or researcher may explore Truth Decay research & commentary to understand the drivers, trends, & consequences of Truth Decay as a System.

PLATFORM/CONTENT GENERATOR TOOLS

1. **SynthID (Digital Watermarking)**
   - **What it does**: Watermarks & identifies AI generated content by embedding digital watermarks directly into AI generated images, audio, text, or video.
   - **How to use it**: Integrated into AI-generated media, detected by compatible tools.
   - **Subscription**: Available via Google Cloud's AI tools (Google DeepMind).
   - **Example**: A fake image of a female CEO is debunked using SynthID detection.

TRANSPARENCY

1. **Hoaxy – Tracking Gendered Disinformation**
   - **What it can do**: Hoaxly visualizes the spread of information online using the X/Twitter & Bluesky APIs.
   - **How to use it**: An API is used to retrieve recent posts matching your search query.
   - **Subscription**: Free until Hoaxly reaches its monthly post limit, then live search is only available to users with Basic ($100/month), Pro ($5000/month), or Enterprise (price available upon request) access.
   - **Example**: Hoaxy reveals bot activity pushing a false claim against a female official.

2. **Systematic Data Collection & Reporting**
   - **What it can do**: Track trends in gendered disinformation & AI-generated attacks.
   - **How it can be used**: Governments, researchers, journalists, & civil society can utilise reports for situational awareness, policy development, & advocacy.
   - **Subscription**: Varies - there is a wide variety of open source reporting available.
   - **Example**: A media watchdog report documents rising deepfake attacks on female politicians, which provides a situational awareness on deepfake trends.

3. **Gender-Sensitive Monitoring**
   - **What it can do**: One can utilise AI tools (e.g. Reality Defender), social network analysis (Hoaxy, Never), & qualitative methods to track gendered disinformation.
   - **How it can be used**: Quantitative / qualitative research to identify gendered attacks online.
   - **Subscription**: Varies - some tools are free, others require paid access.
   - **Example**: Through gender-sensitive monitoring, a researcher is able to show that women candidates face twice as many disinformation attacks as men.

4. **Enhanced User Reporting for Harmful Content**
   - **What it can do**: Improve response time & categorization of gendered disinformation reports.
   - **How it can be used**: As a mass-reporting campaign.
   - **Subscription**: Unknown, would depend on platform implementation.
   - **Example**: A journalist targeted by deepfakes reports it to an enhanced moderation system.

5. **Global Coalition for Digital Safety (World Economic Forum)**
   - **What it can do**: Develop politics & global coordination on digital safety, including gendered disinformation.
   - **How it can be used**: Advocacy groups can engage with the coalition to push for stronger policies.
   - **Subscription**: Dependent on how the coalition is set up.
   - **Example**: An NGO joins the coalition to push for stricter deepfake detection on social media.

Understanding and Countering Gendered Disinformation

Annex E4

Knowledge Resources for Government

To support the development of strategy & briefing materials

CSKA Community Safety Knowledge Alliance

SAPPER LABS

Understanding and Countering Gendered Disinformation

A Framework for Resilience and Action

Knowledge Resources for Government

TO SUPPORT THE DEVELOPMENT OF BRIEFING MATERIALS AND STRATEGIC PLANNING

Purpose

This resource was developed for use by government departments seeking to inform and advise senior public servants and elected officials on options for countering gendered disinformation.

It is intended to serve as a resource, providing an overview of key aspects of gendered disinformation, that may be used in the preparation of briefing materials, strategies and programs.

Source

The source material for this resource is to be found in the following report:

> Botschner, J., Cioffi, G., McMahon, D., Ollinger, J., Sylvestre, B., Kotak, R. & Corley, C. (2025). Understanding and countering gendered disinformation: A framework for resilience and action. Ottawa ON: Community Safety Knowledge Alliance.

This work was prepared by:

CSKA | Community Safety Knowledge Alliance — Research to Practice to Alignment

SAPPER LABS

www.cskacanada.ca www.sapperlabs.com

With funding from:

Canada

Digital Citizenship Contribution Program, Canadian Heritage

The Community Safety Knowledge Alliance (CSKA) is a non-profit applied research organization that supports governments, police, public health and human service leaders in developing, implementing and assessing new approaches to enhancing community safety and well-being service delivery and outcomes. Over the past decade, CSKA has conducted interdisciplinary research and engaged with change-makers on some of Canada's most pressing social issues, including intimate partner violence, youth radicalization to violence, cybersecurity, food security, drug policy, human rights-based policing, and community reintegration initiatives. CSKA maintains an active posture on issues such as disinformation and artificial intelligence to support adaptive responses to these emerging challenges.

Sapper Labs Group (SLG) conducts research to understand the methods and impacts of disinformation and influence campaigns and networks and as input to the development of processes to support effective countermeasures. SLG is supported by global partners and a comprehensive intelligence sharing network. The goal of SLG is make the world a better safer place in line with objectives around: countering foreign interference and influence, countering radicalization and extremism, supporting human rights and other activities involving capacity building related to information integrity.

CONTENTS

Briefing Resource 1: Gendered disinformation as an urgent cross-cutting public policy issue

Addressing Gendered Disinformation: Policy, Legislative, and Research Considerations

ISSUE

Gendered disinformation (GD) is the deliberate spread of false or misleading information targeting women and gender-diverse people, individually, or as a group. Digital technologies – largely social media platforms– are the primary avenues of GD. For this reason, GD may be regarded as a form of technology-facilitated violence against women[1].

Gendered disinformation disproportionately harms women, especially those in politics, media, and activism. Deepfakes and other forms of AI-generated content amplify existing misogyny by sexualizing, discrediting, and silencing women. This form of online violence ultimately leads to:

- Reputation damage – Fake videos or images are used to discredit women in positions of power;
- Psychological and emotional distress – Constant harassment forces many women to self-censor or withdraw from public life;
- Professional consequences – False narratives about women can impact career progression and public trust. It may also discourage other women from progressing their career; or
- Barriers to leadership – Women in leadership or other public positions (e.g. politicians, journalists) face gender-specific attacks that discourage participation in public discourse.

Gendered disinformation may also involve the manipulation of gender as a social wedge to attack women and/or to sway political outcomes[2]. The latter may be used by domestic, ideologically motivated, groups to attack the role of women in Canadian society, as well as by geo-political adversaries as part of foreign interference operations or transnational repression[3].

---

[1] UN Women Expert Group (2022). Technology-facilitated violence against women: Towards a common definition. Report of the meeting of the Expert Group. World Health Organization. https://www.unwomen.org/sites/default/files/2023-03/Expert-Group-Meeting-report-Technology-facilitated-violence-against-women-en.pdf

[2] National Democratic Institute (2022). Interventions for ending online violence against women and girls. https://www.ohchr.org/sites/default/files/documents/issues/expression/cfis/gender-justice/subm-a78288-gendered-disinformation-cso-ndi-annex-3.pdf.

[3] Human Rights Watch. (2024). *We will find you: A global look at how governments repress nationals abroad*. https://www.hrw.org/sites/default/files/media_2024/02/global_transnationalrepression0224web_0.pdf

Online, gender-based, attacks on women in positions of authority (elected officials and others), or those who engage in political advocacy, may be used to try to undermine Canadian democracy. Tactics include actions aimed at undermining the legitimacy of Canadian institutions and/or intimidating women and gender non-conforming individuals into withdrawing from democratic processes.

One in five Canadian women experience some form of harassment on digital platforms. This figure rises to 30 percent of Indigenous women[4]. Between 2019 and 2020, 85 percent of women had witnessed or experienced online gender-based violence and 38 percent had been personally impacted by it[5]. These figures likely under-report the prevalence of the problem.

Gendered disinformation is an escalating public policy issue with implications for:

- The safety of women and girls (both online and offline);
- Online sexual exploitation and technology-facilitated violence (AI-generated deepfakes);
- Hate, bias, and misogyny, including intersections with extremism;
- Foreign interference and digital repression;
- Disproportionate harms to racialized women, particularly Black and Indigenous communities;
- Social cohesion; and
- The vitality and effectiveness of democratic processes.

Female persons are the primary targets of GD. Fake female personas are also used in deception and disinformation campaigns as effective tools for influence or phishing attacks. However, this should not be considered to be a problem of individual women. As a result of its potential for corrosive effects on society, as a whole, GD is a shared, multi-level, threat. It calls for shared responsibility along with joined-up, coordinated action across several levels and sectors of Canadian society. This emerging issue will benefit from coordinated government action in a number of areas to mitigate its harms to the majority of the population. These include: policy, legislation and regulations; research and knowledge mobilization; services and supports; the development of networked capacity; guided by a national framework for action.

[4] Canadian Women's Foundation (n.d.). The facts about gendered digital hate, harassment, and violence. https://canadianwomen.org/the-facts/online-hate-and-cyberviolence/.

[5] Economist Intelligence Unit (2020). Measuring the prevalence of online violence against women. https://onlineviolencewomen.eiu.com/.

BACKGROUND

Gendered disinformation refers to false or misleading digital content used to attack women and gender-diverse individuals, often with the goal of:

- Silencing, discrediting, or harassing women in public life (e.g., politicians, journalists, activists);
- *Ad hominem* attacks on women to discredit, shut down debate or truthful narratives;
- Reinforcing harmful gender stereotypes to discourage women's leadership and participation;
- Undermining survivors of violence through online victim-blaming and misinformation;
- Perpetuating sexual exploitation through synthetic media and deepfake pornography;
- Using gendered disinformation as a weapon of war; and/or
- Spreading extremist narratives that justify violence or discrimination against women.

Gendered disinformation is often spread by:

- Foreign and domestic politically-motivated parties and proxies;
- Religious fundamentalist and/or nationalist movements;
- Misogynistic online communities (e.g., incel groups, men's rights extremists);
- Organized hate groups and extremists using gender narratives to fuel division;
- Foreign state and non-state actors engaged in digital repression and online propaganda; and/or
- Perpetrators of intimate partner violence using online disinformation as a form of coercive control.

In some cases, the identities or affiliations of perpetrators may overlap.

The rate and scope of the spread of gendered disinformation are enabled by the design features and underlying business models of many social media platforms. Examples include the ease with which an item may be re-posted, the algorithms that push content to individual users (in which transparency regarding the reason for particular content appearing is not required to be disclosed or shared), and the interplay between attracting and retaining user attention and platform profitability through targeted advertising revenues.

Adding to the complexity, social media accounts are not verified and posts containing AI-generated content must be voluntarily tagged.

Certain corners of the digital ecosystem, and aspects of the societal environment, create fertile ground for GD to spread and thrive. Factors that may provide the permission structures and conditions conducive to GD include those that: foster social polarization, systems that position female persons in subordinate roles to male persons (patriarchal structures and practices); and beliefs and actions based in mistrust/disdain/hatred of those who are female (misogyny).

PUBLIC SAFETY AND NATIONAL SECURITY IMPLICATIONS

Gendered disinformation positions female persons as both targets and tools of oppression. In some cases, both of these may happen simultaneously. In all cases, female persons – as well as society at-large – are harmed when the safety and participation of those who are female is targeted.

1. Gendered Disinformation as a Threat to Women's Safety Online and in the Physical World

- Technology-facilitated violence: Perpetrators use gendered disinformation, and may leverage free anonymity services and software, and access to dark web resources, to stalk, threaten, or doxx[6] women, especially survivors of violence, racialized women, and LGBTQ2+ individuals.
- Slander and undermined credibility: Fake content can be deployed with the aim of destroying the reputations of female politicians and journalists.[7,8]
- Sexual exploitation and deepfake technology: Increasingly, AI-generated fake sexual content is being used to degrade and extort females.
- Escalation to offline violence: Online disinformation campaign may influence or attempt actively to incite real-world violence, including domestic violence, workplace harassment, and extremist attacks.
- Institutional awareness and response: Law enforcement agencies and social media platforms do not always understand the gravity of, and respond commensurately to, online gendered threats seriously, leaving women with little recourse.[9]

2. Gendered Disinformation and Hate-Based Radicalization

- Connection to hate and extremism: Gendered disinformation fuels radical misogynistic ideologies that intersect with far-right extremism and anti-feminist movements.
- Targeting of racialized women: Black and Indigenous women face intensified online abuse, often shaped by colonial, racist, and sexist narratives.

---

[6] When someone shares private information (such as a home address or phone number) online to intimidate or harm a person.

[7] For example, Chrystia Freeland –manipulated videos promoting fraudulent investment schemes that were circulated online. See https://www.theglobeandmail.com/canada/article-chrystia-freeland-pierre-poilievre-facebook-deepfake-videos-investing/

[8] For example, during election periods, female candidates have reported experiencing coordinated online harassment, (e.g. doctored images and false narratives questioning their competence and integrity). See https://wiisglobal.org/an-overlooked-threat-to-democracy-gendered-disinformation-about-female-politicians/ and https://www.queensu.ca/cidp/publications/policy-briefs/gender-based-disinformation-tool-hybrid-warfare

[9] For example, some have argued that Canada is in need of a cross-sectoral approach to address cyberviolence, suggesting that existing measures were insufficient. See: https://policyoptions.irpp.org/magazines/june-2018/how-cyberviolence-is-threatening-and-silencing-women and https://www.international.gc.ca/world-monde/stories-histoires/2023/2023-03-02-technology-facilitated-gbv-facilitee-technologie.aspx?lang=eng&utm

3. Foreign Interference and Digital Repression

- State-sponsored gendered disinformation to suppress women's voices: Some foreign actors use gendered disinformation to disrupt democratic processes, attempt to suppress the voices of female leaders and destabilize societies.[10,11,12]
- Many wedge issues, that are the target of foreign information manipulation and interference (FIMI), intersect with women's issues.
- Transnational repression: Female human rights defenders and journalists are targeted by foreign-led digital harassment campaigns that seek to undermine their influence by destroying their reputations; these may also extend to real-world physical threats by indirectly influencing ideologically aligned individual violent extremists. [13,14]
- Given the global accessibility. and extra-jurisdictional nature of these platforms, anyone can create an account and target population without any real repercussions.

4. Barriers to Democratic Participation and Public Trust

- Deterring women from leadership roles: Disinformation campaigns disproportionately target women politicians, activists, and journalists, discouraging their public participation.
- Misinformation about women's rights and services: False narratives about feminist organizations, shelters, and reproductive rights mislead the public and hinder policymaking.
- Politicization of women's issues during Canadian elections. [15]

---

[10] For example: Chrystia Freeland has faced persistent online harassment, including doctored images and videos, misogynistic attacks, and threats of violence. In 2022, she was verbally harassed by a man in Alberta who filmed himself aggressively confronting her, using sexist and threatening language. The video was widely shared online, amplifying further attacks against Freeland. For example, see: https://thebreaker.news/news/google-cracks-down-freeland-deepfakes/ https://www.cbc.ca/news/politics/rcmp-investigating-freeland-1.6567275 and https://www.ctvnews.ca/calgary/article/politicians-denounce-video-of-alberta-man-verbally-harassing-deputy-prime-minister-chrystia-freeland/

[11] Online abuse isn't something that only women in the public eye experience—men face threats and hate as well. However, the nature of the harassment women receive is often different. Instead of focusing solely on their work, the attacks frequently target their identity, appearance, or gender. A clear example of this is the treatment of former Environment Minister Catherine McKenna compared to her male successors. McKenna faced relentless gendered abuse, from being mocked as "climate Barbie" to having her campaign office defaced with a misogynistic slur. In contrast, the men who took on the same role after her did not face this kind of gender-specific vitriol. For additional information and examples see https://chatelaine.com/longforms/female-politicians-canada-safety/

[12] For example, see "Why Canadian Politics is Still Unsafe for Female Politicians. https://chatelaine.com/longforms/female-politicians-canada-safety

[13] For example, Chrystia Freeland, as mentioned above (manipulated videos promoting fraudulent investment schemes that were circulated online). See https://www.theglobeandmail.com/canada/article-chrystia-freeland-pierre-poilievre-facebook-deepfake-videos-investing/

[14] For example, during election periods, female candidates have reported experiencing coordinated online harassment, (e.g. doctored images and false narratives questioning their competence and integrity). See https://wiisglobal.org/an-overlooked-threat-to-democracy-gendered-disinformation-about-female-politicians/ and https://www.queensu.ca/cidp/publications/policy-briefs/gender-based-disinformation-tool-hybrid-warfare

[15] A notable Canadian example of gendered disinformation manipulating public opinion involves the online harassment faced by female politicians, which has been shown to have a gendered impact on democratic participation in Canada.

- Destabilizing societies: Foreign actors exploit misogynistic narratives to create division and distrust within a country.[16]

POLICY, LEGISLATIVE and RESEARCH OPTIONS

1. Policy and Legislative Frameworks

- Update hate speech and online harm laws: Expand legal definitions of digital violence to include distribution of non-consensual deepfakes[17] and gendered disinformation as a form of hate-motivated activity. Understand the implications for the right to freedom of expression.[18] Conduct annual assessments to identify risks to women and girls; adjust laws to effectively mitigate risks.[19]
- Mandate that social media platforms label AI-generated content
- Mandate that tools used to create AI-generated content that can be used for exploiting deep fake nudes be banned from app stores and be de-indexed by search engines.
- Strengthen digital rights protections: Implement measures for greater platform accountability for identifying and interrupting the spread of disinformation targeting female persons.[20]
- Enhance gender-sensitive national security strategies: Include and address gendered disinformation as part of foreign interference and domestic extremism threats.
- Pass deep fake abuse laws, including specific attention to AI-generated sexualized content used for gender-based harm.
- Expand the CSIS Act to investigate domestic and foreign information manipulation and interference.
- Appoint a lead government agency to work with the private sector to counter genderized dis-information from domestic and foreign actors.

Research indicates that online harassment fosters a hostile political environment, particularly affecting women, and can influence their willingness to engage in politics. This form of disinformation not only targets individual women but also manipulates public opinion by reinforcing negative stereotypes about women's roles in politics, thereby discouraging female participation and affecting the overall democratic process. For additional information, see https://mlkrook.org/pdf/Wagner_20.pdf

[16] Gendered disinformation has significantly destabilized societies by undermining women's participation in public life and eroding trust in democratic institutions. In Canada, this phenomenon has manifested in various ways, including the spread of misinformation targeting female politicians and activists. For additional information, see https://youtu.be/qZB__f0jU1Y

[17] The European Union has taken significant steps to address the challenges posted by deepfakes through the Artificial Intelligence Act (AI Act), which came into effect in August 2024. The legislation aims to regulate AI technologies - including deepfakes - in order to protect individuals from malicious activities. For additional information, see https://www.bioid.com/2024/06/03/eu-ai-act-deepfake-regulations

[18] For additional information, see https://www.ohchr.org/en/documents/thematic-reports/a78288-gendered-disinformation-and-its-implications-right-freedom

[19] For additional information, see https://www.ohchr.org/en/documents/thematic-reports/a78288-gendered-disinformation-and-its-implications-right-freedom

[20] For example, see https://www.wired.com/story/instagram-gendered-abuse

- Ensure that a future Canadian foreign intelligence service has the necessary authorities to investigate FIMI abroad and to take broad threat reduction measures.
- Work with international partners to streamline the identification of individuals exploiting the platform for genderized disinformation. In addition, create a take-down regime for identified content.
- Ensure that the Canadian Armed Forces and RCMP have the necessary authorities to target gender related on-line violence and disinformation as a weapon of war within sanctioned military operations.
- Revise the definition of Publicly Available Information (PIA) under PIPEDA or equivalent Privacy related legislation to allow for the investigation of disinformation, including end-attribution involving Canadian and foreign entities.

2. Regulatory Measures

- Stronger platform transparency and accountability: Mandate that technology companies report and act on gendered disinformation campaigns.
- Stronger workplace protections: Consider protections for female persons in public-facing roles (e.g., journalists, politicians, academics).
- Examine where elements of the Clean Pipes Strategy[21] might prove beneficial. Recognize that this may entail criticism that it involves government censorship, contrary to interpretations of the value of a free and open speech environment on the internet. Explore options related to reporting and blocking of online gender threats as determined by an independent third party of community members including but not limited to experts, academics, human rights advocates, lawyers, etc. similar to child safety (clean feed) initiatives. Clean Pipes and internet freedom digital rights/net neutrality are not necessarily mutually exclusive. There may be opportunities to harmonize both.

3. Institutional Support

- Governments should offer security assistance for women in public roles. Additionally, independent oversight boards should also be established,[22] and psychological and legal support systems should be made available to vulnerable groups.[23]

4. Research and Funding Priorities

- Expand research and knowledge creation on gendered disinformation: Invest in studies analyzing its impacts on safety, democratic participation, and marginalized groups.

---

[21] This is a cybersecurity approach where internet service providers (ISPs) filter out malicious traffic before it reaches users. This helps protect businesses and individuals from cyber threats like malware, phishing, and denial-of-service attacks by keeping the internet "clean" at the network level.

[22] For additional information, see https://www.ohchr.org/en/documents/thematic-reports/a78288-gendered-disinformation-and-its-implications-right-freedom

[23] For additional information, see https://www.unesco.org/en/articles/fight-against-technology-facilitated-gender-based-violence

- Foster networked capacity for addressing GD as a societal threat: Fund the creation of a cross-sectoral, multi-stakeholder research and knowledge mobilization network.
- Support technology-based solutions: Develop AI tools to act as a shield where it can assist in detecting and combating online gendered disinformation.
- Fund community-led digital safety initiatives: Provide grants to organizations supporting women affected by gendered disinformation.
- Create an OSINT centre-of-excellence and network under a public-private partnership, based on recognition that disinformation is uniquely discoverable as publicly available information (-PAI).

5.Monitoring

- Tune the mandate of the Global Affairs Canada Rapid Response Mechanism (RRM) to also detect, identify and attribute genderized foreign information manipulation targeting Canadian women.
- Fund an industry-supported open source centre-of-excellence to detect, identify and respond to domestic and foreign genderized disinformation.
- Support the development of a sovereign industrial base for technologies, products and services focused on countering dis-information and deep fakes.

RECOMMENDATIONS

Immediate Actions:

- With targeted investment, initiate cross-departmental, industry, academic and private sector operational coordination and program collaboration to address gendered disinformation within public safety, health, digital regulation, defence and national security frameworks.
- Develop a national strategy on gendered disinformation in close partnership with the private sector, integrating public safety, digital governance, and foreign policy approaches.
- Engage women's rights organizations, racial justice groups, security and intelligence professionals and cyber-security experts in the consultation processes. Ensure representation from the governmental, non-governmental/non-profit, academic and private sectors.

Long-Term Priorities:

- Establish gender-responsive online safety laws that hold tech platforms accountable facilitated with a Clean Pipes Strategy.
- Increase data collection and monitoring of gendered disinformation trends and actionable current intelligence.

- Enhance training for security, intelligence, diplomatic, defence, law-enforcement and policymakers on tech-facilitated gender-based violence.

CONCLUSION

Gendered disinformation is a growing community/public safety, digital governance, and national security challenge that disproportionately impacts women, racialized communities, and democracy. A proactive, multi-layered, whole-of-society approach is required to develop policies, legislation, and research agendas that mitigate these harms and safeguard women's rights. Action is needed now to prevent further harm, protect democratic participation, and counter emerging threats, including those involving the interplay between foreign and domestic actors seeking to polarize and disrupt Canadian society and democratic processes.

Briefing Resource 2: Building capacity to tackle gendered disinformation – a cross-sectoral knowledge development and mobilization network

Building capacity to tackle gendered disinformation: A cross-sectoral knowledge development and mobilization network

ISSUE

Addressing gendered disinformation as a public policy issue requires national leadership and a multi-sectoral, coordinated response that brings together platform owners, government, law enforcement, defence, diplomatic, security and intelligence agencies, industry, researchers, civil society, and community organizations. Given the complexity and rapidly evolving nature of gendered disinformation – particularly its intersection with safety, online sexual exploitation, hate-based violence, foreign interference, and its disproportionate impact on marginalized groups, Canada would benefit from the creation of a cross-sectoral knowledge mobilization network – the Gendered Disinformation Knowledge Network (GenD-Net).

Such a network would serve as a hub for leadership, information sharing, research, and policy coordination, program planning, operational coordination and de-confliction ensuring that responses to gendered disinformation are evidence-based, intersectional, and aligned across sectors.

PROPOSED OBJECTIVES OF THE CROSS-SECTORAL KNOWLEDGE MOBILIZATION NETWORK

1. Enhance Knowledge Mobilization & Public Awareness

- Facilitate information and intelligence sharing across government agencies, platform owners, service providers, industry, academia, and civil society organizations.
- Promote public education campaigns on gendered disinformation, digital literacy, and its links to online violence and democratic destabilization.
- Develop and diffuse curriculum resources on gendered disinformation to enhance educational and training opportunities; deliver training to key cross-sectoral actors.
- Ensure that frontline workers, public safety, global affairs, military, intelligence professionals police, and policymakers have access to the latest research and best practices for addressing gendered disinformation.

2. Strengthen Cross-Sectoral Collaboration on Policy Development

- Provide a platform for dialogue between policymakers, industry, researchers, and community organizations.

- Support the co-development of policy recommendations that integrate gender, racial justice, digital governance, and national defence and security considerations.
- Develop guidance documents, legislative proposals, and regulatory strategies based on real-world evidence and lived experiences.

3. Support Defence, Intelligence, Police & Public Safety Agencies

- Develop training programs for police, crown attorneys, national defence, security and intelligence agencies to identify, investigate, and counter gendered disinformation as a form of technology-facilitated violence, coercive control, and foreign interference.
- Provide resources and frameworks for government to collaborate with community organizations in supporting victims.
- Create a national coordination mechanism to track, report, and counter online threats against women in politics, journalism, activism and public life.

4. Advance Research and Innovation

- Establish a research consortium to study the source and impact of gendered disinformation on public safety, extremism, foreign interference, and marginalized communities.
- Support technology-driven solutions (e.g., AI tools for detecting and mitigating gendered disinformation, social media monitoring initiatives, de-indexing sites that promote gendered disinformation).
- Encourage community-led research and participatory approaches, particularly those centering Black, Indigenous, LGBTQ2+, and other disproportionately impacted groups.
- Build evaluation capacity to ensure the effectiveness of interventions focusing on gendered disinformation.

5. Bridge Gaps in Service Provision for Affected Communities

- Connect gender-based violence shelters, digital safety organizations, and legal support networks such as community legal clinics to ensure survivors of online abuse and coercive control receive comprehensive support.
- Provide funding and capacity-building resources to grassroots organizations working on the frontlines of digital safety and disinformation mitigation.
- Ensure that solutions prioritize intersectionality, recognizing that women from racialized and marginalized communities experience gendered disinformation in distinct and compounded ways.

MEMBERSHIP AND GOVERNANCE

The Cross-Sectoral Knowledge Mobilization Network would be composed of representatives from:

- Government Agencies
    - Privy Council Office
    - Public Safety Canada
    - Canadian Armed Forces
    - CSIS (Canadian Security Intelligence Service)
    - RCMP and municipal police forces
    - Global Affairs Canada (to address international dimensions)
    - Canadian Centre for Cyber Security and the Communications Security Establishment
    - Relevant and interested representatives from provincial and territorial governments
- Industry Practitioners and Academic Researchers
    - Experts in intelligence, cyber-security, digital disinformation, radicalization, hate and extremism, gender-based violence, and racial justice
    - Policy institutes and think tanks specializing in disinformation, human rights, democracy, and technology governance
    - Not-for-profit organizations addressing key issues relevant to the public and community safety dimensions of mis- and disinformation
    - Independent digital rights researchers
- Community-Based & Civil Society Organizations
    - Independent fact checking organizations
    - Women's rights organizations
    - Groups representing racialized communities (e.g., Black, Indigenous, and People of Colour-led advocacy groups)
    - Organizations of, and serving, the LGBTQ2+ and gender non-conforming communities
    - Organizations addressing gender-based violence and digital safety

- Technology & Private Sector Stakeholders
    - Representatives from social media platforms, telecommunications and Internet provider firms specializing in AI and analytics, cybersecurity firms, and digital literacy organizations
    - Tech industry leaders working on AI-driven solutions for disinformation and deep fake detection
    - Journalistic organizations working on fact-checking and counter-disinformation efforts

The principal Government of Canada funder could provide secretariate and co-chair functions.

POTENTIAL POLICY AND FUNDING CONSIDERATIONS FOR THE NETWORK

Funding Priorities:

- Core operational funding to establish and sustain the network's activities
- Research grants for projects focused on gendered disinformation, online harms, and foreign interference
- Technology development funds to support organizations developing AI-based detection and mitigation tools
- Capacity-building resources for frontline organizations responding to online threats
- Funding would be accessible to all contributing members of the network.

Policy Levers for Institutional Support:

- Integrate the network's expertise into existing national digital safety, gender equity, and cybersecurity frameworks
- Leverage public-private partnerships to improve industry accountability, transparency and regulatory enforcement
- Strengthen cross-sectoral coordination through dedicated liaison officers within government departments

DEVELOPMENT AND IMPLEMENTATION

- Convene an exploratory roundtable with key stakeholders to assess implementation strategies
- Secure initial funding commitments for research, policy coordination, and knowledge-sharing initiatives
- Develop a roadmap and deliberate action plan for integrating gendered disinformation mitigation into Canada's broader digital governance, national security, cyber security and defence strategies.

CONCLUSION

Gendered disinformation is a growing threat to safety, democracy, and national security – a coordinated response is needed. A knowledge mobilization network would ensure evidence-based policy development and strengthen cross-sectoral collaboration. Such a network would also serve as a critical mechanism for bridging gaps between policy, research, and frontline interventions. Given the complexity of the issue, no single sector can address this challenge alone – a collaboration is essential. Investment in this initiative would enhance Canada's ability to combat online harms, support marginalized communities, and safeguard digital democracy.

Briefing Resource 3: Policy, legislative, and regulatory options

Policy, Legislative, and Regulatory Options for Addressing Gendered Disinformation on Social Media Platforms

ISSUE

Several popular social media platforms are foreign entities usually with roots in America, Russia or China. These foreign social media platforms have become primary vectors for gendered disinformation, both as a form of technology-facilitated violence against women and as a tool of foreign interference and repression. Their algorithmic design, amplification mechanisms, and profit-driven models and leadership contribute to an online environment where misogynistic and harmful content is readily spread, often with minimal accountability. The politicization of these platforms has heightened risks to Canadians. These platforms have historically been reticent to assist Canada in combatting online disinformation.

To effectively mitigate the harms of gendered disinformation, governments must consider a range of policy, legislative, regulatory, incentivizing, and sanctions that target:

1. The amplification and targeting mechanisms of social media platforms
2. The lack of transparency and accountability for harmful content
3. The failure to protect users—particularly women, racialized communities, and LGBTQ2+ individuals—from digital harms

POLICY OPTIONS

Policy options include incentives, regulation, or sanctions focused on social media platforms, in order to support a more adequate capacity and commitment to address gendered disinformation.

1. Regulatory and Legislative Approaches: Strengthening Platform Accountability

A. Establishing a Digital Harms Regulatory Framework

- Mandate platform accountability for online harms by requiring companies to prevent and mitigate the spread of gendered disinformation, much like financial regulations require due diligence in preventing fraud.
- Impose transparency requirements, such as an annual report card on the effectiveness of platforms' moderation of gendered disinformation and content linked to foreign interference.
- Require content removal obligations for gender-based hate speech, deepfake pornography, and known gendered disinformation campaigns by creating and imposing a take-down regime, similar to the process used to remove copyright material and cyber infrastructure.

- Create an independent regulatory body to monitor and audit compliance with online safety obligations.
- Model countermeasure framework off of Clean Pipes Strategy and Child Safety Initiatives.

Example Policy Model:

- The UK's Online Safety Act holds platforms legally responsible for illegal and harmful content, requiring proactive risk assessments and content moderation measures.

B. Algorithmic and Business Model Transparency Requirements

- Require platforms to disclose how their algorithms prioritize, target, and amplify gendered disinformation.
- Mandate external audits on algorithmic biases that disproportionately harm women, racialized communities, and LGBTQ2+ users.
- Limit microtargeting practices that enable misogynistic harassment and disinformation to be directed at specific individuals.
- Implement fairness and transparency standards for content ranking, to prevent the monetization of misogyny and hate-based engagement.
- Mandate a restriction on sponsored material that promotes misogyny and hate-based engagement.

Example Policy Model:

- The EU's Digital Services Act (DSA) requires large platforms to conduct systemic risk assessments on their algorithms and how they contribute to online harms. The DSA also addresses content moderation by requiring platforms to prevent and mitigate against the spread of disinformation, create user protection and impose financial penalties for failing to comply with the requirements.

C. Legal Protections for Victims of Online Gendered Disinformation

- Amend legal frameworks to recognize technology-facilitated gender-based violence (TFGBV) as a specific form of harm.
- Criminalize non-consensual deepfake pornography and establish legal remedies for victims. Currently, with the exception of child sexual abuse material, the Criminal Code of Canada only recognizes a crime when there is a non-consensual sharing of intimate images. If those images are AI-generated, then they do not meet the requirements. (Section 162.1(1))
- . Enhance defamation and privacy tort remedies for individuals harmed by gendered disinformation campaigns
- Expand harassment and hate speech laws to explicitly include disinformation-driven campaigns targeting women and marginalized groups.

Example Policy Model:

- Canada's Bill C-36 (proposed) sought to amend hate speech provisions to better address online harms, including gender-based hate.

  Canada's Online Harms Act (Bill C-63), officially titled, *An Act to enact the Online Harms Act, to amend the Criminal Code, the Canadian Human Rights Act and An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service and to make consequential and related amendments to other Acts*, aimed to address harmful content on the internet. In particular, issues related to child exploitation, hate speech, and content promoting violence or self-harm. The Bill would establish a Digital Safety Commission to oversee compliance, investigate complaints, and enforce penalties. The Bill also aims to hold platforms accountable for the content that is hosted on their platform. In particular, it creates several Duties on the platform such as a duty to act responsibly, protect children and keep all the records. If the Bill were to receive Royal Assent, then the legislation would increase penalties for hate crime, expand the definition of hate crime and amend elements of the Criminal Code of Canada. It needs to be noted that the Bill was not passed prior to the 2025 election, hence, it is currently not codified in law.

2. Incentivizing Compliance: Economic and Market-Based Strategies

A. Financial Penalties for Failure to Address Gendered Disinformation

- Impose heavy fines on platforms that fail to act on gendered disinformation or that allow repeated violations of their terms of service and/or community standards.
- Create a tiered penalty structure—higher fines for platforms with greater reach and influence.
- Tie penalties to the volume and severity of violations, as determined by independent oversight bodies.

Example Policy Model:

- Under the EU's Digital Services Act, platforms that fail to remove harmful content can face fines of up to 6% of global annual revenue.

B. Tax Incentives for Platforms that Invest in Safer Online Environments

- Offer tax credits/reductions/exemptions for companies that proactively implement gender-sensitive AI moderation and content detection.
- Encourage investment in trust and safety teams by providing financial incentives for companies that exceed minimum moderation standards.

Example Policy Model:

- Some governments have experimented with green tax incentives—a similar model could be applied to digital safety compliance.

3. Requiring Greater Transparency and Reporting from Platforms

A. Mandatory Disclosure of Disinformation Campaigns

- Require platforms to disclose information about coordinated gendered disinformation campaigns, particularly those linked to foreign actors.
- Establish real-time reporting obligations when gendered disinformation is being used to harass public figures, journalists, or activists.

Example Policy Model:

- The US Honest Ads Act (proposed) aims to increase transparency on political advertising and misinformation on platforms.
- Manitoba has proposed a bill to address deepfakes and election disinformation which addresses the use of AI and other tools to mislead voters[24].
- BC has legislation to prosecute those who use AI generated deep fakes to interfere with an election or mislead voters.

B. Independent Auditing and Whistleblower Protections

- Mandate external audits to ensure platforms are upholding commitments to tackle gendered disinformation.
- Create whistleblower protections for former employees who expose platform failures.

Example Policy Model:

- The Facebook whistleblower revelations demonstrated how internal research confirmed that the platform amplified harmful content, underscoring the need for external accountability.

4. Addressing Gendered Disinformation as a National Security Concern

A. Strengthening Counter-Disinformation Mechanisms

- Develop a national strategy to combat gendered disinformation as part of Canada's cybersecurity, defence policy and national security priorities.
- Increase funding for research on and countermeasures against foreign-led gendered disinformation campaigns and their impact on democratic institutions.
- Enhance inter-agency coordination between Public Safety Canada, CSIS, DND, CSE, GAC the RCMP, and international allies.

[24] https://www.thecanadianpressnews.ca/politics/manitoba-bill-takes-aim-at-deepfakes-election-disinformation-voter-misdirection/article_13551407-bfc7-5e63-b3ff-29c6d781d469.html

B. Holding Social Media Platforms Accountable for Foreign Interference

- Legislate mandatory cooperation with Defence, Global Affairs, as well as intelligence and security agencies when platforms are used for foreign disinformation operations.
- Impose restrictions on foreign-controlled accounts and influence networks that spread misogynistic disinformation.

Example Policy Model:

- The EU's Anti-Disinformation Code of Practice requires platforms to monitor and disrupt foreign information operations.

RECOMMENDATIONS

- Conduct cross-sector consultations with experts in gender-based violence, cybersecurity, open source intelligence, national security, and digital regulation.
- Develop a regulatory framework that ensures platform accountability, transparency, and meaningful financial penalties for non-compliance.
- Invest in digital literacy, research funding, open source intelligence and enforcement mechanisms to strengthen Canada's resilience against gendered disinformation.

CONCLUSION and RECOMMENDATIONS

The design and business models of social media platforms incentivize the spread of gendered disinformation by rewarding engagement over safety. In some cases, amplifying harmful content enhances profitability. Without stronger legislative, regulatory, and enforcement measures, platforms will continue to prioritize corporate revenue over public safety. Action is required to mitigate the impacts of the role that these platforms have with respect to the scale of negative impacts of gendered disinformation with respect to social cohesion, domestic stability and well-being and national security.

Briefing Resource 4: Gendered disinformation as a national security threat

Gendered Disinformation as a National Security Threat

ISSUE

Gendered disinformation – the deliberate spread of false or misleading information targeting women and gender-diverse individuals – is not only a domestic issue related to the safety of women and girls. It is also increasingly a tool of foreign interference, digitally-enabled transnational repression, and asymmetric hybrid warfare used by geopolitical adversaries to:

- Undermine democratic institutions and national stability by discrediting women in leadership, politics, media, and civil society.
- Suppress dissent and opposition through online harassment and coercion, targeting women activists, journalists, and public figures.
- Weaken allied coalitions and institutions by amplifying misogynistic narratives that deepen societal divisions.
- Directly threaten, intimidate and harm over half the Canadian demographic.
- Use online platforms to execute disinformation campaigns that disproportionately affect marginalized communities, including Black, Indigenous, and LGBTQIA+ individuals.
- Target socio-political wedge issues which are predominantly women's issues, to create division in Canadian society

Addressing gendered disinformation requires an enhanced national security posture, including new operational policies, regulations, diplomatic responses, and military doctrines that recognize this as an element of modern warfare and digital repression.

We must recognize and highlight the targetable vulnerabilities associated with gender-centric disinformation on our democratic institutions, public health and safety, national defence, cyber security, global affairs and the rule of law.

Additionally, cross-sectoral collaboration – involving governmental, non-governmental, and private-sector actors – will be crucial in detecting, countering, and mitigating these foreign-led operations.

BACKGROUND

The Use of Gendered Disinformation in Foreign Interference and Digital Repression

Women are disproportionately targeted with mis-dis-mal-information (MDM) particularly using AI generated deep fakes. Women in positions of power are even more exposed to this type of warfare.

Gender-related MDM and transnational repression have significant national security and defense implications, as they can discredit leadership, exacerbate gender divides, influencing political outcomes, destabilize societies, undermine democratic institutions, stall programs, erode public trust and become a vector for cyber-attacks. Targeting female leadership in Canada with disinformation and deep fakes presents a number of national security threats:

Undermining public trust: By spreading false information, manipulating media or generating deep fakes against female leaders, adversaries seek to undermine public confidence in these figures and, by extension, leadership and the democratic institutions and businesses they represent. A loss of trust in government officials, or industrial enterprises can distract critical operations, delay important programs, destabilize governance, lead to societal unrest or adversely affect markets.

Influencing political outcomes: Disinformation campaigns can negatively impact election outcomes or policy decisions by swaying public opinion based on false premises. Targeting female leaders may exploit existing social biases, potentially discouraging support for progressive policies or diminishing their electoral prospects.

Exacerbating gender divides: Such tactics can deepen existing gender divides, fostering environments of discrimination and exclusion. This polarization can weaken social cohesion and create divisions that are exploitable by hostile entities. Many wedge issues exploited by hostile actors using manipulative information campaigns are predominately women's issues.

Hindering diversity and representation: By casting doubt on the credibility of female leaders – claiming they only achieved the position of power owing to affirmative diversity, equity, and inclusion (DEI) – is designed to de-legitimize their position and hinder efforts to achieve greater female representation in leadership roles. This can impact decision-making processes, stalling the advancement of important national programs and policies.

Damaging international reputation: Attacks on female leaders can tarnish Canada's global image as a progressive and inclusive society. This could affect diplomatic relations and partnerships, impacting trade, security cooperation, and global influence.

Attack vectors: Salacious fake stories involving female leadership or celebrities constitutes a popular bait-and-click (phishing) attack vector for cyber exploitation and further disinformation amplification.

State-Sponsored Disinformation Campaigns Targeting Women in Politics, Journalism, and Activism

- Russia, China, India, Iran, USA (recently) and other state actors have engaged in targeted online campaigns to discredit and harass female politicians, journalists, and human rights defenders in Western democracies.
- Tactics include deepfake pornography, false narratives about personal lives, and accusations of corruption, instability, and/or incompetence.
- Example: Russian-backed campaigns have targeted female leaders in NATO-aligned states, seeking to erode public trust in their leadership.

Gendered Disinformation as a Tool of Hybrid Warfare

- Foreign adversaries weaponize gendered narratives to divide societies, increase polarization, and degrade trust in democratic institutions.
- Example: Disinformation campaigns have amplified "anti-feminist" rhetoric to stoke resentment and extremism, weakening domestic social cohesion.

Online Repression of Women's Rights Advocates and Journalists

- Authoritarian regimes suppress dissent through targeted gendered disinformation attacks on women's rights activists and journalists reporting on state abuses.
- Example: China's use of online campaigns to discredit Uyghur women who speak out against human rights abuses, or Iran's repression of female activists following mass protests.

The Role of Emerging Technologies in Amplifying Gendered Disinformation

- Artificial intelligence (AI) and deepfake technology, also known as "nudify" applications, have enabled the creation of non-consensual sexualized content to discredit women in leadership and civil society.
- Much of this software is created for the porn industry, stored extra-jurisdictionally off-shore and used by a variety of bad actors including state proxies, child predators and political supporters.
- Social media algorithms reward outrage and engagement, ensuring misogynistic disinformation spreads widely with minimal accountability.

Implications

These Tactics Techniques and Procedures (TTPs)[25] pose a direct national security risk, requiring enhanced threat monitoring, operational countermeasures, and policy frameworks to defend against them.

OPERATIONAL and POLICY RECOMMENDATIONS FOR SECURITY, INTELLIGENCE, AND MILITARY COMMUNITIES

1. Strengthening Intelligence and Security Responses to Gendered Disinformation

- Develop a security and intelligence doctrine on gendered disinformation as a foreign threat
    - Recognize gendered disinformation as a vector of foreign interference in national security assessments.
    - Incorporate gendered disinformation into threat modeling, cybersecurity protocols, and intelligence-sharing frameworks.
- Expand cybersecurity and counter-disinformation operations
    - Enhance cyber-threat detection to identify state-sponsored gendered disinformation campaigns.
    - Implement automated and AI-driven tracking of foreign-facilitated harassment campaigns targeting women in leadership roles.
- Strengthen protections for women in public service and military leadership
    - Establish security protocols for online threats against female government officials, military personnel, and diplomats.
    - Enhance legal and policy protections for individuals targeted by foreign-led disinformation attacks.
    - Enhance the civil remedies available to victims of foreign-led disinformation attacks.

  Comment: Defensive cyber operations against gendered disinformation may be supported by a proposed integration of two operational structures – the DISARM Framework and the

[25] TTPS describe how cyber threats operate:

- Tactics – The overall goals or objectives of an attack (e.g., gaining access, stealing data).
- Techniques – The specific methods used to achieve the goals (e.g., phishing, exploiting vulnerabilities).
- Procedures – The detailed steps or actions taken to execute the techniques (e.g., sending a fake email with a malicious link).

F3EAD Intelligence Loop within proactive defensive cyber operations targeting influence operation cyber kill chain. The DISARM Framework[26] "provides a common language and structure for defenders to share data and analysis, and coordinate whole of society responses to malign influence operations." The DISARM process has been endorsed by a number of ally organizations, including the NATO/EU European Centre of Excellence for Countering Hybrid Threats and the EU cybersecurity agency, ENISA. The F3EAD[27] intelligence loop[28] was developed for counterterrorism, military and special forces to enable the fusion of intelligence and military operations when targeting high-value adversaries. Increasingly used within cyber threat intelligence, F3EAD is used to remove or restrict adversaries from information ecosystems. A proposed integration of these two approaches is outline in Attachment A.

2. Enhancing Diplomatic and International Security Strategies

- Integrate gendered disinformation into foreign policy and diplomatic engagements
    - Address gendered disinformation as a violation of international norms and digital human rights in global cybersecurity dialogues.
    - Work with allied nations, NATO, and multilateral bodies to strengthen international frameworks on online repression.
    - Expand Global Affairs' Rapid Response Mechanism (RRM) to detect and report gendered disinformation.
- Impose diplomatic and economic consequences on perpetrators
    - Implement sanctions and diplomatic responses against states engaging in gendered disinformation campaigns.
    - Develop trade and technology policies that restrict the export of AI-based disinformation tools to repressive regimes.

3. Funding and Building a Cross-Sectoral Network to Counter Gendered Disinformation

A coordinated response involving government, private-sector intelligence and technology firms, civil society organizations, and academic institutions is necessary to:

- Improve private-public intelligence-sharing on gendered disinformation campaigns;
- Enhance the mandate and capacity for the Canadian government to proactively target and disrupt domestic and foreign disinformation infrastructures and actors;

[26] https://www.disarm.foundation/
[27] F3EAD stands for the sequence: Find, Fix, Finish, Exploit, Analyze and Disseminate
[28] https://kravensecurity.com/f3ead-loop/

- Facilitate governments ability to outsource intelligence and operational support services to detect, attribute, target and counter gendered dis-information;
- Update government policy to allow private sector to securely share intelligence and provide secure means of communicating this intelligence at scale and speed;
- Develop a national gendered disinformation threat landscape reporting capacity which could, in-turn, feed into a publicly accessible intelligence "dashboard" to support awareness and collaborative action (Attachment B);
- Enhance detection and active countermeasures within defence, intelligence cyber-security and law enforcement agencies;
- Develop advanced AI-driven countermeasures to mitigate online foreign interference and repression; and
- Fund research into the long-term security implications of gendered disinformation.

Establishing a cross-sectoral network focusing on fusing and mobilizing knowledge, and *coordinating action*, to address gendered disinformation would serve as a hub for the development of pan-Canadian networked capacity.

Key Stakeholders for the Network:

- Security & Intelligence Agencies: Privy Council Office, Public Safety, Canadian Security Intelligence Service, Communications Security Establishment, National Defence, Global Affairs, Royal Canadian Mounted Police.
- Allied National Security Bodies: NATO StratCom, Five Eyes intelligence-sharing partners.
- Telecoms, ISPs, platform providers, major IT and Data companies.
- Industry - Intelligence and defence industries, social media firms, cybersecurity companies, AI developers.
- Academic & non-academic research groups: Cybersecurity, AI, disinformation and public/community safety experts.
- Civil society & women's organizations: Ensuring an intersectional approach to countering online repression.

RECOMMENDATIONS

- Establish a dedicated government funding stream (contract vehicle) for research, innovation and policy development on gendered disinformation. Open to Canadian industry, academia and not-for profit organizations.
- Incentivize Canadian industry participation and innovation through public-private partnerships and direct investment.

- Develop a national strategy on gendered disinformation as a foreign interference threat, and ensure integration with national defence policy, cyber security and national security strategies.
- Fund the creation of a cross-sectoral intelligence-sharing network to combat gendered disinformation.
- Establish legal and policy frameworks to protect women in public life from foreign-backed digital threats.
- Develop a rapid response mechanism to protect individuals facing high-risk foreign disinformation attacks.

CONCLUSION

Gendered disinformation is not just a gender equality issue—it is a national security threat. Governments have an opportunity to frame and appropriate and effective fusion of defensive and offensive options by integrating gendered disinformation into:

- National security doctrines and intelligence frameworks;
- Cyber threat detection and counter-disinformation strategies; and
- Diplomatic and military engagement on foreign interference and hybrid warfare.

Additional information with references for further reading is included in Attachment C.

Attachment A: Integrated Framework for Use in Defensive Cyber Operations

Proposed integration of F3EAD and DISARM frameworks within proactive defensive cyber operations targeting influence operation cyber kill chain

Overview of F3EAD and DISARM frameworks

- F3EAD is used by special operations forces (SOF) to anticipate and predict enemy operations, identify, locate, and target enemy forces, and to perform intelligence exploitation and analysis in a rapid cyclic manner in a constantly changing environment.
  - F3EAD involves six elements:
    - Find: Locate the adversarial assets or targets.
    - Fix: Confirm and track the target to ensure precise engagement.
    - Finish: Neutralize the target through direct action.
    - Exploit: Gather intelligence from the operation for further use.
    - Analyze: Assess collected intelligence to refine future operations.
    - Disseminate: Share insights with stakeholders to enhance collective effectiveness.

- DISARM is a methodology designed to counter disinformation and influence operations. It emphasizes the systematic identification and neutralization of disinformation narratives.
  - In one form, it consists of five elements:
    - Detect: Identify disinformation campaigns and emerging narratives
    - Interpret: Understand the content, intent, and potential impact of the narrative
    - Segment: Analyze and categorize the target audience or affected demographics
    - Analyze: Assess the spread, actors, and ecosystem supporting the disinformation
    - Respond: Develop counter-messages or strategic actions to disrupt the disinformation
    - Mitigate: Implement measures to prevent future occurrences
  - Another version uses a 'Red Team (mock threat actors)/Blue Team' (mock defenders) approach to calibrate mitigation measures against various scenarios involving malicious tactics, techniques and procedures

Integration of F3EAD and DISARM frameworks and application to influence operation cyber kill chain

Stage 1: Target Identification
DISARM: Detect & Interpret
➢ Use social media analytics, sentiment analysis, and network analysis to detect hostile narratives or campaigns.
DISARM 2: Plan strategy, plan objectives
➢ Blue Framework – responder TTPs
F3EAD: Find
➢ Identify the key actors, influencers, and channels spreading disinformation.

Stage 2: Target Fixing
DISARM: Segment
➢ Segment the audience impacted by the disinformation and identify vulnerable groups.
DISARM 2: Micro-target; develop content; select channels and affordances; establish social assets
DISARM Blue Framework – responder TTPs
F3EAD: Fix
➢ Confirm the identities of key disinformation actors and map their influence networks.

Stage 3: Engagement and Neutralization
DISARM: Respond
➢ Deploy counter-narratives, fact-checking, and media campaigns to disrupt the influence of disinformation.
DISARM 2: Conduct pump priming; deliver content; persist in the information environment
DISARM Blue Framework – responder TTPs
F3EAD: Finish
➢ Conduct targeted actions to neutralize adversarial actors, such as de-platforming accounts or discrediting fake sources.

Stage 4: Intelligence Exploitation
DISARM: Analyze
➢ Examine the effectiveness of countermeasures and assess the persistence of disinformation.
DISARM 2: Assess effectiveness
DISARM Blue Framework – responder TTPs
F3EAD: Exploit
➢ Gather and process intelligence from neutralized actors, including their communication methods, tactics, and goals.

Stage 5: Continuous Adaptation
DISARM: Mitigate
➢ Implement broader security measures to prevent future influence campaigns, such as public education and platform policies.
DISARM 2: Adjust tactics as necessary, and re-initiate the cycle as required
DISARM Blue Framework – responder TTPs
F3EAD: Analyze & Disseminate
➢ Share operational insights with allied agencies and organizations to refine global counter-disinformation strategies.

DISARM framework: Definitions

Source: https://disarmframework.herokuapp.com/

Plan strategy: "Define the desired end state, i.e. the set of required conditions that defines achievement of all objectives."

Plan objectives: "Set clearly defined, measurable, and achievable objectives. Achieving objectives ties execution of tactical tasks to reaching the desired end state. There are four primary considerations: - Each desired effect should link directly to one or more objectives - The effect should be measurable - The objective statement should not specify the way and means of accomplishment - The effect should be distinguishable from the objective it supports as a condition for success, not as another objective or task."

Micro-target: "Target very specific populations of people."

Develop content: "Create or acquire text, images, and other content."

Select channels and affordances: "Selecting platforms and affordances assesses which online or offline platforms and their associated affordances maximize an influence operation's ability to reach its target audience. To select the most appropriate platform(s), an operation may assess the technological affordances including platform algorithms, terms of service, permitted content types, or other attributes that determine platform usability and accessibility. Selecting platforms includes both choosing platforms on which the operation will publish its own content and platforms on which the operation will attempt to restrict adversarial content."

Conduct pump priming: "Release content on a targeted small scale, prior to general release, including releasing seed. Used for preparation before broader release, and as message honing. Used for preparation before broader release, and as message honing."

Deliver content: "Release content to general public or larger population."

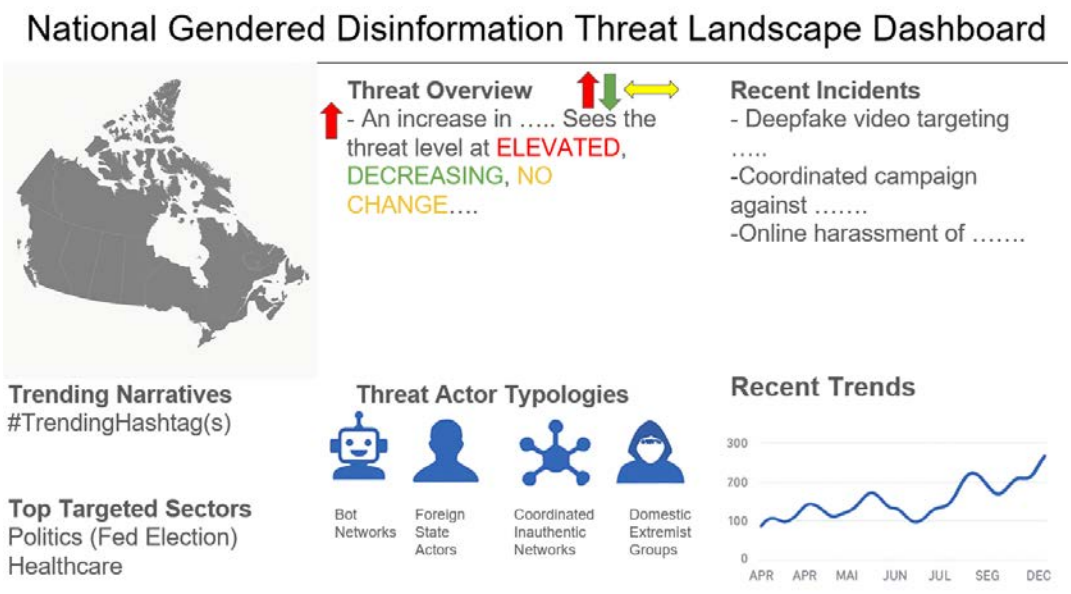
Persist in the information environment: "Persist in the Information Space refers to taking measures that allow an operation to maintain its presence and avoid takedown by an external entity. Techniques in Persist in the Information Space help campaigns operate without detection and appear legitimate to the target audience and platform monitoring services. Influence operations on social media often persist online by varying the type of information assets and platforms used throughout the campaign."

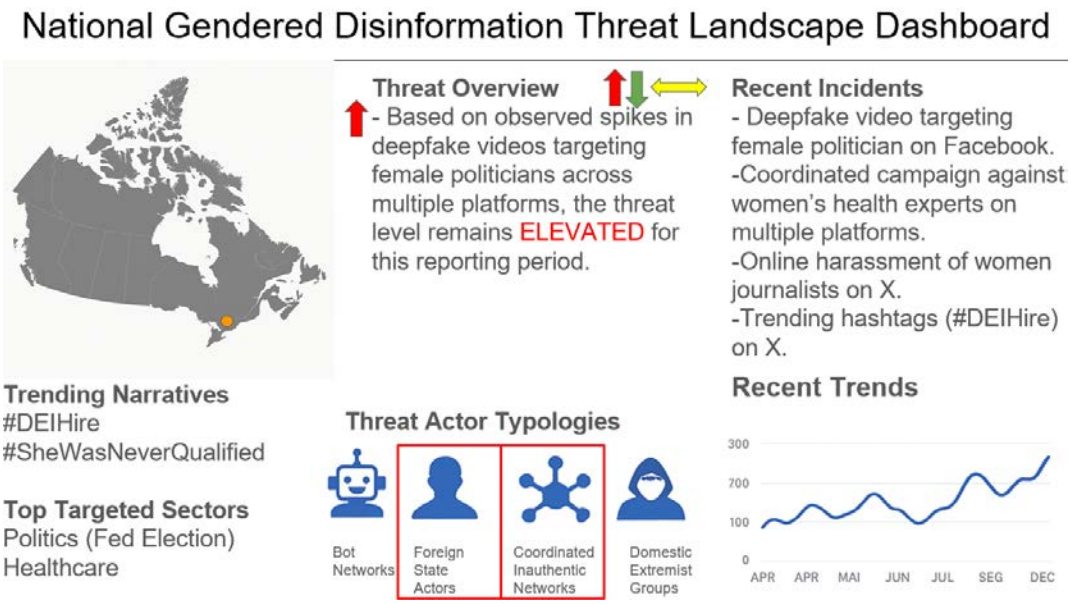
Assess effectiveness: "Assess effectiveness of action, for use in future plans."

Establish social assets: "Establishing information assets generates messaging tools, including social media accounts, operation personnel, and organizations, including directly and indirectly managed assets. For assets under their direct control, the operation can add, change, or remove these assets at will. Establishing information assets allows an influence operation to promote messaging directly to the target audience without navigating through external entities. Many online influence operations create or compromise social media accounts as a primary vector of information dissemination."

Attachment B: Sample Depiction of Gendered Disinformation Threat Landscape Dashboard

(Data are hypothetical)





Attachment C: Additional Information and References for Further Reading

Technology-Facilitated Gender-Based Violence as a Multi-Systemic Issue

Technology-facilitated gender based violence (TFGBV) occurs across multiple interconnected systems:

- Individual: Women, girls, and LGBTQI+ individuals experience disproportionate impacts (Quilt.AI & ICRW, 2021). Those with intersecting marginalized identities face increased vulnerability (Sambasivan et al., 2019).
- Relationship: Perpetrators are often known to victims, including intimate partners, family members, and colleagues (Hassan et al., 2018; Koirala, 2020).
- Community: Schools and workplaces often lack adequate support mechanisms and awareness (Gurumurthy et al., 2019). Cultural norms and taboos around gender and sexuality contribute to victim-blaming (Chowdhury, 2016).
- Societal: Patriarchal structures and gender inequalities are replicated and amplified online (Gurumurthy et al., 2019). Legal frameworks are often insufficient or poorly enforced (Halder, 2017).
- Technological: Platform design and policies can enable abuse, while anonymity emboldens perpetrators (Dunn, 2020). The digital divide leaves some users more vulnerable (UN Women, 2020).

TFGBV is deeply embedded in existing social structures and power dynamics. Effective interventions must therefore target multiple levels simultaneously, from individual digital literacy to societal norm change. The interconnected nature of these systems also means that progress (or lack thereof) in one area can have ripple effects across others.

Chowdhury, S. (2016). Report of expert consultation responding to violence against women and girls in the cyber age. CARE Bangladesh. https://www.researchgate.net/publication/319269676_Report_of_Expert_Consultation_Responding_to_Violence_against_Women_and_Girls_in_the_Cyber_Age.

Dunn, S. (2020). Technology-facilitated gender-based violence: An overview. *Centre for International Governance Innovation: Supporting a Safer Internet Paper No. 1.*, https://ssrn.com/abstract=3772042.

Gurumurthy, A., Vasudevan, A., & Chami, N. (2019). Born Digital, Born Free?: A Socio-Legal Study on Young Women's Experiences of Online Violence in South India. IT for Change. DOI: 10.13140/RG.2.2.14962.94407.

Hassan, B., Unwin, T., & Gardezi, A. (2018). Understanding the Darker Side of ICTs: Gender, Sexual Harassment, and Mobile Devices in Pakistan. *Information Technologies & International Development*, 14, 1-17. https://www.researchgate.net/publication/323737497_Understanding_the_Darker_Side_of_ICTs_Gender_Sexual_Harassment_and_Mobile_Devices_in_Pakistan.

Koirala, S. (2020). Female journalists' experience of online harassment: A case study of nepal. *Media and Communication, 8*(1), 47-56. https://doi.org/10.17645/mac.v8i1.2541.
Quilt.AI and ICRW. (2021). COVID-19 and online violence in India: Digital intelligence report. https://www.icrw.org/wp-content/uploads/2021/10/Online-Violence_Full-Report-Quilt.AI-and-ICRW.pdf.

Sambasivan, N., Consolvo, S., Batool, A., Ahmed, N., Matthews, T., Thomas, K., Gaytán-Lugo, L. S., Nemer, D., Bursztein, E., & Churchill, E. (2019). They don't leave us alone anywhere we go. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. https://doi.org/10.1145/3290605.3300232.

UN Women. (2020). *Online and ICT* facilitated violence against women and girls during COVID-19*. https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2020/Brief-Online-and-ICT-facilitated-violence-against-women-and-girls-during-COVID-19-en.pdf.

Gender as a Target of Harm

The prevalence of deepfake pornography targeting women represents a serious violation of privacy, consent, and bodily autonomy. This technology enables a new form of image-based sexual abuse that can have devastating psychological, social, and professional consequences for victims (Bates, 2017). The ease of creating and distributing deepfakes allows individual bad actors to weaponize this technology against women on a concerning scale, perpetuating misogyny and the objectification of women's bodies online. The affective dimensions of gendered deepfakes are particularly harmful. They can cause severe emotional distress, reputational damage, and have a silencing effect on women's participation in public life. The threat of deepfakes may discourage women from seeking positions of leadership or speaking out on important issues. The hyperrealistic nature of deepfakes can sow doubt about genuine content, allowing perpetrators to claim real evidence of misconduct is fake, further complicating efforts to combat gendered disinformation campaigns.

Aikenhead, M. (2021). Revenge pornography and rape culture in canada's nonconsensual distribution case law. in J. Bailey, A. Flynn & N. Henry (Eds.) *The emerald international handbook of technology-facilitated violence and abuse*. Leeds, UK: Emerald. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3945880.

Bates, S. (2017). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn on female survivors. Feminist Criminology, 12(1), 22-42.

Burgess, M. (2020). A deepfake porn bot is being used to abuse thousands of women. WIRED.

Chisala-Tempelhoff, S., & Kirya, M. T. (2016). Gender, law and revenge porn in Sub-Saharan Africa: A review of Malawi and Uganda. *Palgrave Communications, 2*, 16069. https://www.nature.com/articles/palcomms201669.

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.

Powell, A., Henry, N., & Flynn, A. (2018). Image-based sexual abuse. In W. S. DeKeseredy & M.

Wang, C. (2019). Deepfakes, revenge porn, and the impact on women. *Forbes*, November 1, 2019. https://www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and-the-impact-on-women/.

Gender as a Tool of Harm

The use of deepfakes to target women represents a significant threat to women's participation in public life and democratic processes. By leveraging gender stereotypes and sexualized content, state actors can effectively silence and discredit female voices. This technology exacerbates existing online harassment issues faced by women, potentially discouraging their involvement in politics and activism. The creation of false female personas also allows state actors to manipulate public opinion and spread disinformation more effectively, exploiting societal trust in certain female archetypes.

Cohen, Z., Lyngaas, S., & Perez, E. (2024, May 15). Exclusive: US intelligence spotted Chinese, Iranian deepfakes in 2020 aimed at influencing US voters. *CNN*. https://www.cnn.com/2024/05/15/politics/us-intelligence-china-iran-deepfakes-2020-election/index.html.

Lee, S. F. (2024, March 29). Deepfakes with Chinese characteristics: PRC influence operations in 2024. *China Brief, 24*(7), 21-28.

Tactics, Techniques and Procedures

The proliferation of deepfake technology and gendered disinformation poses significant threats to national security, democratic processes, and individual rights. These tactics can erode public trust, manipulate elections, and silence marginalized voices, particularly women in public life. The low barrier to entry and potential for rapid, widespread dissemination make these threats particularly challenging to combat, requiring coordinated efforts from governments, tech companies, and civil society to develop effective countermeasures.

These sophisticated tactics and techniques represent a significant escalation in the capabilities of foreign actors to conduct influence operations. The ability to create highly convincing fake content at scale poses unprecedented challenges for democratic societies, media literacy, and national security. The combination of AI-powered content generation with targeted dissemination strategies makes these campaigns particularly insidious and difficult to counter.

As deepfake technology continues to advance, it will become increasingly challenging for the public and even experts to distinguish between real and synthetic content. This erosion of trust in visual and audio evidence could have far-reaching implications for public discourse, journalism, and the functioning of democratic institutions.

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace.

Busch, E., & Ware, J. (2023). *The Weaponisation of Deepfakes: Digital Deception by the Far-Right*. International Centre for Counter-Terrorism.

Farooq, N. (2024). *Content Warfare: Combating Generative AI Influence Operations*. Tech Policy Press. https://www.techpolicy.press/content-warfare-combating-generative-ai-influence-operations/.

Lee, S. F. (2024). Deepfakes with Chinese Characteristics: PRC Influence Operations in 2024. *China Brief, 24*(7), 21-28.

Milmo, D. (2024). Russia targets Paris Olympics with deepfake Tom Cruise video. *The Guardian*. June 3, 2024. https://www.theguardian.com/technology/article/2024/jun/03/russia-paris-olympics-deepfake-tom-cruise-video.

Smith, H., & Mansted, K. (2020). *Weaponised deep fakes: National security and democracy*. Australian Strategic Policy Institute.

United States Department of State. (2023). *Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors*. https://2021-2025.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/.

Threat Assessment

The threat of deepfakes is rapidly evolving and becoming more accessible to malicious actors - state and non-state alike. This technology has the potential to erode trust in digital media, manipulate public opinion, and cause significant harm to individuals and organizations. The disproportionate targeting of women and vulnerable groups is particularly concerning. As deepfake technology becomes more sophisticated and widely available, it will likely exacerbate existing social and political divisions, posing a serious challenge to democratic institutions and social cohesion.

Canadian Security Intelligence Service (CSIS). (2023). *The evolution of disinformation: A deepfake future*. https://www.canada.ca/en/security-intelligence-service/corporate/publications/the-evolution-of-disinformation-a-deepfake-future.html.

Department of Homeland Security (DHS). (2021*). Increasing threat of deepfake identities*. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

Kira, B. (2024). *Deepfakes, the weaponisation of AI against women and possible solutions*. https://verfassungsblog.de/deepfakes-ncid-ai-regulation/.

How Gendered Disinformation Achieves Its Effects.

Gendered disinformation weaponizes gender stereotypes and sexist narratives to undermine the credibility and competence of women in politics (Judson, 2021; Sessa, 2022). It often combines true information with misleading presentation, unprovable rumors, or value judgments to attack targets in ways that are not straightforwardly false (Judson, 2021); it also exploits social media algorithms and platform designs that tend to amplify extreme, negative, and polarizing content

(Sessa, 2022; Wilfore, 2021). Gendered disinformation uses coordinated inauthentic behavior, including bots and troll networks, to artificially amplify messages (Di Meco & Wilfore, 2021). Gendered disinformation taps into existing prejudices and stereotypes to evoke emotional responses and gain credibility with certain audiences (Judson, 2021). The combined effect of these tactics is to create a hostile online environment for women in politics, discourage women's participation in public life, and ultimately undermine democratic processes and consolidate power for illiberal actors (Di Meco & Wilfore, 2021; Sessa, 2022). By exploiting technological and social vulnerabilities, gendered disinformation campaigns can have outsized impacts beyond their factual basis, posing threats to both gender equality and broader democratic values.

Di Meco, L., & Wilfore, K. (2021). *Gendered disinformation is a national security problem*. Brookings Institution. https://www.brookings.edu/articles/gendered-disinformation-is-a-national-security-problem/.

Judson, E. (2021). *Gendered disinformation: 6 reasons why liberal democracies need to respond to this threat.* Heinrich-Böll-Stiftung European Union. https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat.

Sessa, M. G. (2022). *What is gendered disinformation?* Heinrich-Böll-Stiftung European Union. https://il.boell.org/en/2022/01/26/what-gendered-disinformation.

Wilfore, K. (2021). The gendered disinformation playbook in Germany is a warning for Europe. Brookings Institution. https://www.brookings.edu/articles/the-gendered-disinformation-playbook-in-germany-is-a-warning-for-europe/.

Countermeasures

Effective responses require a multi-stakeholder approach involving civil society, platforms, political actors, and election authorities. Integrating gender analysis into broader counter-disinformation efforts is crucial to address the disproportionate impacts on women and marginalized groups.

EU DisinfoLab. (2021). *Gender based disinformation: Advancing our understanding and response. https*://www.disinfo.eu/publications/gender-based-disinformation-advancing-our-understanding-and-response/.

His Majesty the King in Right of Canada (2024). Countering disinformation: A guidebook for public servants. https://www.canada.ca/content/dam/di-id/images/pcdi/guide-eng.pdf.

Judson, E., Atay, A., Krasodomski-Jones, A., Lasko-Skinner, R., & Smith, J. (2020). *Engendering hate: The contours of state-aligned gendered disinformation online*. Demos.

This page is intentionally left blank

CSKA Community Safety Knowledge Alliance | SAPPER LABS