Input for Special Rapporteur's Annual Thematic Report on Disinformation
Nina Jankowicz, Disinformation Fellow
Woodrow Wilson International Center for Scholars
Washington, DC, USA
February 2021

I appreciate the opportunity to contribute to the Special Rapporteur's work on disinformation. Over the past eight years, through my roles as an implementer of democracy support programs, a Fulbright Fellow and strategic communications advisor to Ukraine's Foreign Ministry, and a leading researcher and chronicler of information operations in Central and Eastern Europe, the United States, and beyond, I have witnessed online disinformation develop from a curiosity of the modern era to a threat that presents a mounting challenge to free expression and democratic participation. I have been privileged to witness this evolution from some of the places where this clash comes into starkest relief.

In an effort to keep this submission brief, I will provide an overview of the key trends I have observed in my work since January 2020, as well as worrisome regulatory and private sector responses that I have analyzed in that time, with suggestions for further, more detailed reading in footnotes.

Gendered and Sexualized Disinformation
While conducting qualitative interviews on disinformation in Central and Eastern Europe in 2017, I came across several examples of state-backed information operations targeting prominent women's gender and sexuality in order to undermine their credibility, fitness for office, and public support.[1] In reporting on this phenomenon, I coined the term "sexualized disinformation." Other researchers began to explore this theme.[2] The Wilson Center and Moonshot CVE undertook a mixed methods research project in late 2020 to define, quantify, and evaluate the use of online gendered and sexualized disinformation campaigns against women in public life, as well as inform efforts to address the phenomenon in the public and private sectors.[3]

Tracking 13 women politicians across six social media platforms for two politically tumultuous months, we uncovered over 330,000 instances of gendered and sexualized abuse and disinformation. We noted the widespread use of what we dubbed "malign creativity," or the use of coded language; iterative, context-based visual and textual memes; and other tactics to avoid detection on social media platforms. We found that, social media platforms are not doing nearly enough to address the problem; thanks to malign creativity and platform structures created by cisgender white men for users like themselves, platforms and content moderation policies often fail to imagine the myriad and unique attacks which abusers employ against women and people of color in the public sphere, as well as the effect these attacks have on their targets.

---

[1] Nina Jankowicz, "How Disinformation Became a New Threat to Women," *Coda Story*, 11 December 2017.
[2] See: Samantha Bradshaw, "The Gender Dimensions of Foreign Influence Operations," Global Affairs Canada, 2019; Lucina Di Meco, "#SHEPERSISTED: Women, Politics & Power in the New Media World," The Wilson Center, Fall 2019; Ellen Judson, Asli Atay, Alex Krasodomski-Jones, Rose Lasko-Skinner, and Josh Smith, "Engendering Hate: The contours of state-aligned gendered disinformation online," Demos, October 2020.
[3] Nina Jankowicz et al, "Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online," The Wilson Center, 25 January 2020.

Using two focus groups and three in-depth interviews, the report examines the toll that gendered and sexualized disinformation campaigns have on women in public life, and how little recourse women feel they have in reporting structures available to them. It also underlines that belligerent and antidemocratic states weaponize online and offline misogyny to create compelling influence campaigns in their own domestic and foreign environments. Overall, the continued use of these tactics, met with little resistance from platforms and policymakers and described as "the cost of doing business" for women in public life threatens women's equal participation in conversations on the internet and off. The effects of such campaigns are broader than just the target in question; for every incident in which gendered and sexualized narratives against a high-profile female target are allowed to proliferate, influencing the target's public presence, thousands of other women see those narratives and consider whether to engage at all.

We make several recommendations for policymakers, platforms, and employers in the report. The most pressing are:

- ***Social media platforms*** should introduce incident reports that allow women to report multiple abusive posts at once to provide more context and a more holistic view of the abuse they are experiencing.
- They should also regularly update platform classifiers or keywords to reflect and root out malign creativity, improve automated detection methods, and introduce nudges or friction to discourage users from posting abusive content.
- Finally, they should create a cross-platform consortium to track and respond to online misogyny, similar to existing consortiums which counter terrorism and extremism.

- ***Lawmakers*** should include content moderation transparency reporting requirements in social media regulation bills to improve understanding of the problem and introduce accountability for women's online protection.
- They also should create clear standards that prohibit the use of gendered and sexualized insults and disinformation in official business.

Increased dependence on closed spaces as vectors of disinformation
For the past several years, Facebook users have been seeing more content from "friends and family" and less from brands and media outlets.[4] As part of the platform's "pivot to privacy" after the 2016 election, groups have been promoted as trusted spaces that create communities around shared interests. "Many people prefer the intimacy of communicating one-on-one or with just a few friends," explained Mark Zuckerberg in a 2019 blog post. "People are more cautious of having a permanent record of what they've shared."

But as research my colleague Cindy Otis and I conducted shows, those same features—privacy and community—are often exploited by bad actors, foreign and domestic, to spread false information and conspiracies. Dynamics in groups often mirror those of peer-to-peer messaging apps: People share, spread, and receive information directly to and from their closest contacts, whom they typically see as

---

[4] This section adapted from: Nina Jankowicz and Cindy Otis, "Facebook Groups are Destroying America," *WIRED*, 17 June 2020.

reliable sources. To make things easier for those looking to stoke political division, groups provide a menu of potential targets organized by issue and even location; bad actors can create fake profiles or personas tailored to the interests of the audiences they intend to infiltrate. This allows them to seed their own content in a group and also to repurpose its content for use on other platforms.[5]

As investigations from *The Wall Street Journal*,[6] *The Markup*,[7] and other outlets have shown, Facebook is aware of the problems groups pose, but has done little to fix them. The changes the platform has made -- altering recommendation algorithms, increasing oversight in parts of some political groups -- have happened too late. Groups containing threats to public safety and public health continue to proliferate on the platform with little awareness or moderation, as the January 6 events at the US Capitol demonstrated.[8] Communities with hundreds of thousands of members exist in near secrecy and share and amplify content in direct contravention to Facebook's Community Standards. Some of the most urgent changes to address these challenges include:

- Facebook should entirely eliminate group recommendations for *all* groups, as we find that some groups that contain civic and health content are willfully mislabeled.
- Private and secret groups should be capped at a certain number of members, and undergo regular human-led content moderation for adherence to terms of service.
- Lawmakers should introduce transparency regulations to better understand how groups affect information consumption and user behavior, as well as the platform's content moderation decisions.

The same dynamics are at play in the closed messaging apps to which millions have flocked in the past several months. Telegram's "channels," WhatsApp's "broadcast lists" and groups, and Signal's groups all pose similar challenges to Facebook groups: closed environments, lack of moderation and oversight. Unlike on Facebook, however, messaging platforms do not use algorithmic recommendations to drive engagement or group membership, and some platforms cap the number of users that can be added to a group or that a message can be forwarded. Civil society groups like First Draft News have led efforts to counter messenger disinformation by establishing a tipline on the platform itself, where users can send spurious messages.[9] Platforms should consider creating, advertising, and training users in endemic reporting functions that could mimic such tiplines.

Regulatory responses
As more countries wake up to the threat of online disinformation, regulation looms, and with it, threats to free expression, both inadvertent and intentional. While platforms' self-regulation is untenable, so too are regulatory solutions that empower political bodies to crack down on rivals. Even governments with benign motivations behind their counter disinformation regulations, such as Germany and Brazil, have run

[5] Nina Jankowicz, "How an Anti-Shutdown Celebrity Is Made," *The Atlantic*, 3 October 2020.
[6] Jeff Horwitz, "Facebook Knew Calls for Violence Plagued 'Groups,' Now Plans Overhaul," *The Wall Street Journal*, 31 January 2021.
[7] Leon Yin and Alfred Ng, "Facebook Said It Would Stop Pushing Users to Join Partisan Political Groups. It Didn't.," *The Markup*, 19 January 2021.
[8] Jane Lytvynenko, ""Stop The Steal" Groups Are Still Flourishing On Facebook," *BuzzFeed News*, 8 January 2021.
[9] Pedro Burgos, "What 100,000 WhatsApp messages reveal about misinformation in Brazil," First Draft News, 27 June 2019.

up against the dangerous normalization of restrictions on speech. Germany's NetzDG law has been cited in authoritarian countries as fodder for their own crackdowns on freedom of expression. Brazil, the country that passed the world's first online bill of civil rights, flirted with passing harsh fines for the spread of disinformation. Conversely, Singapore and Ukraine have both used the threat of disinformation to excuse anti-democratic actions against political opposition and dissenting opinions. In a cross-continental, cross-cultural, and cross-contextual examination of these four approaches to regulation, I distilled five guiding principles for any regulation aimed at countering online disinformation while protecting democratic ideals:[10]

- First, when defining what speech will be deemed harmful in counter disinformation regulation, precision is key. In both the Singaporean and Ukrainian cases, overbroad definitions contributed to fears that laws drafted under the guise of protecting national security and freedom of opinion would rather contribute to a chilling effect on free speech as well as empower the government to quash criticism.
- Second, the case studies demonstrate the importance of mandating transparency and oversight—ideally from an apolitical, expert body—as part of any regulatory framework.
- Third, the importance of establishing an independent body to enforce and adjudicate counter disinformation law, ideally drawing on the existing structures and expertise of judicial authorities, cannot be understated. Any body overseeing these laws should be expert, politically insulated, and utilize the independent judiciary for adjudication.
- Fourth, users must have recourse above the platform level in order to dispute takedowns of their content. They must be informed of its removal as well as of the opportunities they have to challenge the decision. These appeals should move through the aforementioned independent, expert commissions charged with overseeing and enforcing social media regulation.
- Finally, the development of any social media regulation should be pursued in consultation with civil society and other democratic partners, and with the use of existing legal frameworks.

Too often, discussions about responding to disinformation and preserving freedom of expression are presented in an unhelpful binary, as if the only options are an information free-for-all or government censorship. As we have seen amid the coronavirus pandemic, online disinformation has harmful offline effects on public safety and public health. The United Nations and the Special Rapporteur should suggest frameworks and guiding principles for regulation and responses that preserve transparency, empower individuals, and underline the idea that the rampant spread of disinformation has offline consequences and silencing effects, in particular for women and minorities.

---

[10] Nina Jankowicz and Shannon Pierson, "Freedom and Fakes: A Comparative Exploration of Countering Disinformation and Protecting Free Expression," The Wilson Center, December 2020.