

DEMOS

# ENGENDERING HATE

THE CONTOURS OF  
STATE-ALIGNED GENDERED  
DISINFORMATION ONLINE

ELLEN JUDSON

ASLI ATAY

ALEX KRASODOMSKI-JONES

ROSE LASKO-SKINNER

JOSH SMITH

OCTOBER 2020

## Open Access. Some rights reserved.

Open Access. Some rights reserved. As the publisher of this work, Demos wants to encourage the circulation of our work as widely as possible while retaining the copyright. We therefore have an open access policy which enables anyone to access our content online without charge. Anyone can download, save, perform or distribute this work in any format, including translation, without written permission. This is subject to the terms of the Creative Commons By Share Alike licence. The main conditions are:

- Demos and the author(s) are credited including our web address **www.demos.co.uk**
- If you use our work, you share the results under a similar licence

A full copy of the licence can be found at **<https://creativecommons.org/licenses/by-sa/3.0/legalcode>**

You are welcome to ask for permission to use this work for purposes other than those covered by the licence. Demos gratefully acknowledges the work of Creative Commons in inspiring our approach to copyright. To find out more go to **[www.creativecommons.org](http://www.creativecommons.org)**



This study was commissioned by NDI



Published by Demos October 2020

© Demos. Some rights reserved.

15 Whitehall, London, SW1A 2DD

T: 020 3878 3955

[hello@demos.co.uk](mailto:hello@demos.co.uk)

[www.demos.co.uk](http://www.demos.co.uk)

# CONTENTS

|   |                |
|---|----------------|
| <b>ACKNOWLEDGEMENTS</b>                                   | <b>PAGE 4</b>  |
| <b>EXECUTIVE SUMMARY</b>                                  | <b>PAGE 5</b>  |
| <b>FINDINGS AND RECOMMENDATIONS</b>                       | <b>PAGE 7</b>  |
| <b>INTRODUCTION</b>                                       | <b>PAGE 10</b> |
| <b>CHAPTER 1</b><br>UNDERSTANDING GENDERED DISINFORMATION | <b>PAGE 11</b> |
| <b>CHAPTER 2</b><br>BACKGROUND                            | <b>PAGE 16</b> |
| <b>CHAPTER 3</b><br>THE PLAYERS                           | <b>PAGE 19</b> |
| <b>CHAPTER 4</b><br>THE RULES OF THE GAME                 | <b>PAGE 25</b> |
| <b>CHAPTER 5</b><br>TALKING BACK AND SPEAKING UP          | <b>PAGE 33</b> |
| <b>CHAPTER 6</b><br>THE GENDERED DISINFORMATION ECOSYSTEM | <b>PAGE 35</b> |
| <b>CHAPTER 7</b><br>CONCLUSIONS                           | <b>PAGE 37</b> |
| <b>CHAPTER 8</b><br>RECLAIMING THE GAME                   | <b>PAGE 39</b> |

# ACKNOWLEDGEMENTS

First and foremost, huge thanks must go to NDI, without whose support, collaboration and generosity of time and expertise this report would not have been possible. In particular, thanks must go to Molly Middlehurst, who I have had the pleasure to work closely with over the last year in developing the methods and preparing this report, and whose keen eye and contributions at the later stages of the report were crucial. Thanks also to Sandra Pepera, Tamar Eisen, Caroline Hubbard, and Moira Whelan for their invaluable support and contributions to this work.

We owe a great debt of gratitude to the in-country experts in Poland and in the Philippines, who laid the foundations that made this research possible, through the preparation of the political and lexical analysis we used to guide our research, and through their continued guidance.

Thanks to all the participants in the closed roundtable held to discuss the findings of this report, and all those in government, media, and civil society with whom we've engaged on this crucial topic.

And of course, thanks to our Demos colleagues, past and present, for all their hard work and support. Thanks must go to Maeve Thompson, for her constant support and commitment to ensuring the impact of this research; Josh Tapper for launching this report; Bibi Nubir for looking after the financials; Shivpreet Singh for his proofreading skills; and Ava Berry, Sonny Edwards and Elliot Jones for their contributions early on in the project.

Any and all mistakes or omissions are our own.

Ellen Judson  
Asli Atay  
Alex Krasodonski-Jones  
Rose Lasko-Skinner  
Josh Smith

October 2020

# EXECUTIVE SUMMARY

'Attacks on women which use hateful language, rumour and gendered stereotypes combine personal attacks with political motivations, making online spaces dangerous places for women to speak out.'

Online spaces are being systematically weaponised to exclude women leaders and to undermine the role of women in public life. Attacks on women which use hateful language, rumour and gendered stereotypes combine personal attacks with political motivations, making online spaces dangerous places for women to speak out. And left unchecked, this phenomenon of gendered disinformation, spread by state and non-state actors, poses a serious threat to women's equal political participation.

In this research, we investigated state-aligned gendered disinformation in two countries, Poland and the Philippines, through an analysis of Twitter data. We analysed tweets in Polish and, from the Philippines, in English. All Twitter data was collected through the platform's public API. Data was collected and analysed as follows:

1. A list of users who heavily influence online political discourse in each country was assembled, and tweets sent by these accounts and a sample of users following them was collected.
2. Tweets were filtered to those containing terms within a lexicon of political and gendered terminology, which was built by in-country experts and refined according to language discovered in the dataset.
3. The resulting dataset was filtered to identify tweets containing a name or phrase associated with a likely target of gendered disinformation, based on initial analysis and input from in-country experts.
4. This data was manually analysed to identify which tweets displayed gendered disinformation, and the key themes which arose.
5. Finally, the dataset was used to generate network maps of which users were sharing gendered disinformation, and whether there was evidence of state-alignment in those who were most engaged in this activity.

In August 2020, a closed roundtable was held by NDI with experts from across the world to discuss the findings and the ramifications for mitigation strategies.

What did we find?

- Gendered disinformation is being shared by state-aligned actors online - though it reaches much broader audiences.
- The way in which gendered disinformation spreads across a network varies hugely according to context.
- But the themes of gendered disinformation - the rules that it follows - are often broadly consistent.
- Gendered disinformation is parasitic on news events, existing rumours, and underlying social stereotypes, and seeks to reshape the terms of political discourse in a way that harms women.
- Gendered disinformation plays on existing tropes to try to convince people that women in public life, are one or more of: devious, stupid, overly sexual, in need of protection, or immoral: and so unfit for public life.

- Gendered disinformation is not just false information - it uses highly emotive and value-laden content to try to undermine its targets.
- Gendered disinformation weaponises harassment against women in public life, and tries to make them afraid to talk back.
- But women are speaking up: online counterspeech is being used by women in public life and their allies to fight back.

What does this mean for the fight against disinformation?

- Gendered disinformation should not be overlooked in responses to general disinformation. Solutions in policy and practice must recognise that disinformation takes many forms, and can vary what it looks like and where it originates according to context.
- Centralised automated solutions to gendered disinformation are likely to censor legitimate speech and overlook gendered disinformation. Input and oversight from local experts who understand the language and context in which disinformation occurs is vital.
- Solutions should centre and learn from the experiences of women who are already working to challenge disinformation: the problem is a systemic one and targets of gendered disinformation should not be expected to fix it as individuals.

# FINDINGS AND RECOMMENDATIONS

## WHAT IS STATE-ALIGNED GENDERED DISINFORMATION?

Online spaces are being used to weaponise gendered language and narratives about what people of different genders should and should not do in order to pursue political goals. In this report, we focus on gendered disinformation as it affects women: not only as it is likely to disproportionately affect women, but because of the specific role that gendered disinformation plays in the systematic exclusion of women from political power.

‘Gendered disinformation’ refers to information activities (creating, sharing, disseminating content) which:

- Attacks or undermines people on the basis of their gender;
- Weaponises gendered narratives to promote political, social or economic objectives.

Gendered disinformation online exists at the intersection of disinformation with online violence, such as abuse and harassment: it seeks impact primarily at the political level, though can also cause serious harm at the personal level.

This online behaviour, as with online violence against women in politics, operates on several levels to achieve political outcomes: targeting individual women to harm them or drive them out of public life, while also sending a message that women in general should not be involved in politics: that it is too risky, too dangerous, too unsuitable, too detrimental to a state.<sup>1</sup> This message is intended for other women who are in politics or considering it; but also for the demos

at large. By playing on existing stereotypes and tropes about women, underlying beliefs that women simply are not good enough for politics are reinforced, with the ultimate aim that women self-censor and fail to garner support from voters or citizens at large.<sup>2</sup>

We sought in this research to investigate state-aligned gendered disinformation (SAGD), whereby actors who are part of a state, or whose behaviour or interests align with those of a state, engage in gendered disinformation to promote political outcomes. Gendered disinformation can be deliberate, without being coordinated. We use the broad term ‘state-aligned’ to include disinformation created and shared not only by those who may have been directly coordinated by the government, but also by informal networks of actors who use gendered disinformation to try to shield the state from critique or democratic threat, in particular by women in public life.

Such state-aligned gendered disinformation has been identified and can be seen in political situations across the world: especially that which “mixes old ingrained sexist attitudes with the anonymity and reach of social media in an effort to destroy women’s reputations and push them out of public life”.<sup>3</sup> These campaigns take multiple forms, but with common gendered elements: from doctored images claiming to be naked photos of women politicians who have spoken out against war to fake photos of women journalists which purport to show they have had plastic surgery.<sup>4,5</sup> Other examples include screenshots purporting to be from women politicians’ sex tapes, harassment campaigns by troll armies, and misleading or

1. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

2. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

3. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

4. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

5. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 2020]

fabricated content used against prominent women of colour who criticise the state leader.<sup>6,7,8,9</sup>

## KEY FINDINGS:

We found gendered disinformation being shared more often within certain clusters of online Twitter users in Poland and the Philippines: notably clusters of politically influential people, including those who have alignment in some way with the state.

However, gendered disinformation does not always take place within an echo chamber. We found that where gendered disinformation is shared in a network is very context-dependent: in Poland, it was shared at a low level amongst a more interconnected network. In the Philippines, gendered disinformation was concentrated more in one group which had less overlap with other groups of users.

Counterspeech and gendered disinformation are often shared by the same groups of people, though this speech often takes different, politically-inflected forms in different groups.

Through engaging in online political discourse in set ways, whether that is the use of certain words, hashtags, tropes or targeting particular people and groups, the rules of online political discourse itself are re-shaped, into a language game which excludes women as full, active and respected members.

It is through this lens that we present six 'rules': these are patterns we consistently observed in this research in how women in public life were attacked and spoken about, and gendered narratives which were being manipulated by state-aligned actors and their networks. We focused on Poland and the Philippines specifically, but these rules have echoes globally in the kinds of abuse women are subjected to and the narratives which authoritarian leaders

**'These rules have echoes globally in the kinds of abuse women are subjected to and the narratives which authoritarian leaders aim to promote to consolidate their own power and political agendas.'**

aim to promote to consolidate their own power and political agendas.

These rules are:

### RULE 1

#### **Convince others that women are devious: they are not fit for politics**

A central narrative in gendered disinformation is that women are not fit for public or private political lives, and as such, should be excluded. Most frequently, this language surrounds their character: as presented through gendered disinformation, women are devious, unreliable and fake.

### RULE 2

#### **Denounce women as too stupid for public life**

Through a mix of cited pseudoscience, old-fashioned 'humour' and general abuse, the stupidity of women was a regular topic of gendered disinformation in the datasets we examined.

### RULE 3

#### **Make women afraid to talk back**

Over the past five years, platforms have made changes aimed at improving the experience of vulnerable users, though targeting of women in coordinated online abuse continues today. Women calling out this harassment, or using tools to reduce it through blocking or muting accounts, was often mentioned in abuse. On top of this, platform decisions seen to be protecting women were also decried.

### RULE 4

#### **Praise women for being sexy, condemn them for being sexual**

Sexualised distortion is common in disinformation campaigns against women - from sharing doctored photos of women politicians naked, to screenshots purporting to be from sex tapes, to accusations of illicit affairs. The power of these tactics comes from social norms which pervade about women and their sexuality - from the 'virgin' archetype that women should remain pure, to the 'whore' - that women are obscenely sexual and will use their sexuality to get what they want.

6. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

7. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 2020]

8. Posetti, J. Fighting back against prolific online harassment in the Philippines. The Conversation, 2017. Available at <https://theconversation.com/fighting-back-against-prolific-online-harassment-in-the-philippines-80271> [accessed 03 May 2020]

9. Holmes, J. A Cop's Threat Against Alexandria Ocasio-Cortez Shows Propaganda Is Dangerous. Esquire, 2019. Available at <https://www.esquire.com/news-politics/a28479868/alexandria-ocasio-cortez-police-officer-threat-misinformation/> [accessed 03 May 2020]



## **RULE 5:**

### **Show everyone that the strong men will save them**

The character of strongmen in politics is making a comeback. These 'strong men' demonstrate similarities in their attitudes towards women. When women try to reclaim their power or raise their voice, they see it as a challenge to their rule as well. Hence they encourage an undermining discourse towards women and other genders, to support their policies such as banning abortion, their anti-LGBT sentiments and their ideas on where women belong.

## **RULE 6:**

### **Demonise the values that women hold**

The amalgamation of Left-wing politics and feminist politics was a central feature of gendered disinformation explored by this research. By eliding the two, conservative or Right-wing governments such as those investigated portray feminism as anti-governmental and oppositional.

Overall, we found that the ways that those who engage in gendered disinformation act can be characterised as follows:

**THEY USE EMOTIONS, NOT FACTS:** The clearest commonality between the stories told about women in public life described above is that they seek to inflame people's emotions, in addition to spreading false information about their targets.

**THEY TARGET WOMEN, CONVERT THEIR ALLIES, BOLSTER THEIR CRITICS:** The messages work to engender anger, disgust and disdain in the third party reader; and fear and shame in the second person target.

**THEY USE IN-JOKES:** Specific phrases or terms are used in online discourse by communities engaging in SAGD to denigrate a particular woman or make reference to a particular event, which might not be immediately obvious to an outside observer.

**THEY MAKE THE NEWS - THEN USE THE NEWS:** Abuse or offensive comments made by political leaders become a discussion point in themselves which lead to repetition online

**THEY ARE PARASITIC:** Disinformation relies on co-opting underlying values, beliefs and ways of speaking to get its message across, rather than trying to invent something completely new.

## **RECOMMENDATIONS**

We need systematic improvements to be made to how platforms operate in facilitating, promoting and moderating online speech, and how policy and practice respond to these problems. We present the challenges of existing solutions to these problems, and criteria that effective solutions must meet.

### **RECOGNISE THE DIVERSITY OF**

**DISINFORMATION:** Disinformation takes many forms and is engaged in and perpetuated by many different actors: a 'one-size fits-all' approach will not be successful. Fact-checking, while important in the fight against disinformation, is not a panacea, as emotional content and value-judgements play a major role in gendered disinformation. Similarly, focusing on formal networks risks overlooking the informal networks that are implicated in gendered disinformation: the political influencers and their followings.

**NOT JUST A 'GENDER' PROBLEM:** Tackling gendered disinformation should not be the sole purview of gender-based programming: broader responses to online harms, analysis of information operations, or improvements to platform design and oversight, should take into account the ramifications for the spread or encouragement of gendered disinformation.

**CENTRE LOCAL EXPERTISE:** How gendered disinformation is framed and spreads across a network varies greatly according to context. Platforms should support the work of local experts in identifying and combating gendered disinformation, for instance through the provision of data access or the trialling of potential responses through changes to platform design. Automated systems for identifying gendered disinformation are unlikely to have high levels of accuracy - though if employed, should be employed transparently and overseen by local experts.

**ELEVATE WOMEN'S VOICES:** Solutions should be co-created with people who have experienced being the targets of gendered disinformation, to ensure that the most significant threats are being addressed in a way that does not compound the problem.

# INTRODUCTION

In a post-truth era worries about disinformation have become commonplace: from claims of “fake news” levelled at everyone from blogs and social media accounts to mainstream media organisations, to the mainstreaming of conspiracy theorists and the exploitation of online spaces by state and non-state actors. The common conception of what it even is for something to be true is under attack.

But disinformation does not affect everyone equally. Women in public life are on the frontlines facing the weaponisation of information spaces against them. Online threats, libellous claims, doctored explicit images, vitriolic abuse and harassment: all attempts to disempower women, bolster support for political attacks on them, and destabilize their critiques, particularly of powerful men.

This report investigates the rules that state-aligned actors are following in their attacks on women in public life; and how women are fighting back. We examine the content that is being shared; the platforms which are involved; and the actors who engage in gendered disinformation. We also present a methodology, co-created with the National Democratic Institute (NDI) - which commissioned this study - for examining state-aligned gendered disinformation (SAGD) online.

First, we take a deep-dive into the concept of ‘gendered disinformation’, examining what it means and how it relates to disinformation more generally and harassment. We build on work by the NDI and others to understand the ways in which gender norms are manipulated and utilized in disinformation campaigns to influence political views and outcomes. (*Understanding Gendered Disinformation*, page 11).

We then present the background for this research, including context around gender and politics in the two countries we have looked at for this report, Poland and the Philippines (*Background*, page 16).

We also characterise the users involved - those who are engaging in sharing content relevant to gendered disinformation, and where state-alignment is present in the relevant online actors, through network maps of how they relate to each other (*The Players*, page 19).

We outline six Rules: the methods and aims of gendered disinformation found in our study, and characterise the ways they use content to reduce women’s participation in political, public and social life (*The Rules of the Game*, page 25). We present our conclusions about the overarching aims and methods of gendered disinformation (*Conclusions*, page 37).

We examine how other platforms are connected to or reflect the gendered disinformation we were seeing on Twitter, and examine their profiles and relative influence online of relevant users (*The Gendered Disinformation Ecosystem*, page 35).

We then examine counterspeech in the Philippines as a particular instance of where women and others are calling out these attempts and speaking up against harassment and abuse (*Talking Back and Speaking Up*, page 33).

Finally, we discuss the ramifications for responding to gendered disinformation, and outline the criteria which successful responses must meet. (*Reclaiming the Game*, page 39)

We lay out our full network analysis and methodology in the *Appendices* (see separate document). We also consider the successes and limitations of the methods we used, and how they could be refined in future (*Limitations, Challenges and Methodology Recommendations*).

# CHAPTER 1

# UNDERSTANDING

# GENDERED

# DISINFORMATION

## WHAT IS GENDERED DISINFORMATION?

Actors engage in information operations, such as disinformation, when they coordinate to covertly use information insincerely to achieve political, social or economic goals.<sup>10</sup>

Our understanding of ‘information’ is broad, and includes content which may not make clear assertions but nevertheless communicates or expresses claims. As we showed in our report *Warring Songs*, disinformation can include emotive or value-laden content rather than merely fact claims. We identified four main strategies used by information operators: to change what information is seen by others; to obfuscate the spaces in which information is shared; to engender sympathy for their cause; and to limit the participation of opposing voices in these spaces. Central to online disinformation campaigns is the weaponisation of online spaces for political purposes - not necessarily only the proliferation of falsities.<sup>11</sup>

Here we explore a form which information operations can and do take, which demonstrate these strategies - that is, using online spaces to weaponise gendered language and narratives about what people of different genders should and should not do in order to pursue political goals. In this report, we focus on gendered disinformation as it affects women: not only as it is likely to disproportionately affect women, but because of the specific role that gendered disinformation plays

‘Gendered disinformation online exists at the intersection of disinformation with online violence, such as abuse and harassment: it seeks impact primarily at the political level, though can also cause serious harm at the personal level. Essentially it consists in the weaponisation of rumour and stereotypes: with false, misleading or hateful narratives told, often in abusive language, in order to achieve a political impact.’

in the systematic exclusion of women from political power.

Gendered disinformation online exists at the intersection of disinformation with online violence, such as abuse and harassment: it seeks impact primarily at the political level, though can also cause serious harm at the personal level. Essentially it consists in the weaponisation of rumour and stereotypes: with false, misleading or hateful narratives told, often in abusive language, in order to achieve a political impact.

Gendered disinformation, as with violence against women online, is also likely to affect different groups of women in different ways, and “[target] one or more aspects of a woman’s identity

10. Krasodomski-Jones, A. and others. *Warring Songs: Information Operations in the Digital Age*. Demos, 2019, pp. 11-13. Available at <https://demos.co.uk/wp-content/uploads/2019/05/Warring-Songs-final-1.pdf> [accessed 03 May 2020]

11. Thanks to participants at a closed roundtable for their valuable discussion of this topic.

(e.g., racism, transphobia, etc.)”, overlapping, for instance, with racist disinformation when women of colour are the targets.<sup>12</sup>

Forms of “information disorder” online are often categorised into discrete categories - such as disinformation, where false information is deliberately created and spread; misinformation, where false or misleading information spreads, and malinformation, where harassment or abuse is used for political purposes.<sup>13</sup> All of these categories can be identified in gendered content online - in particular, the extent of gendered harassment and abuse online is well documented.<sup>14</sup> Work by the National Democratic Institute (NDI) found that online violence against women in politics could be categorised in a way analogous to violence against women in politics (VAW-P), as involving: insults and hate speech; embarrassment and reputational risk; physical threats; and sexualised distortion. These categories were used as the foundation for some of our analysis for this report.<sup>15</sup>

However, the complexity of gendered information online is such that frequently gendered harassment and abuse cannot be separated from the promotion of false information about its target; that gendered misinformation, even if spread without the knowledge that it is false, can also be abusive and have political intentions and contribute to a toxic online environment; and that fabricated information gains traction by appealing to pre-existing gendered norms to become common parlance. As such, drawing these tight distinctions risks excluding the complexity of gendered information online.

As such, we use ‘gendered disinformation’ as an umbrella term and present the following working definition:

**‘Gendered disinformation’ refers to information activities (creating, sharing, disseminating content) which:**

- Attacks or undermines people on the basis of their gender;

- Weaponises gendered narratives to promote political, social or economic objectives.

Gendered disinformation should be understood as a pattern, much like violence against women in politics more generally. “One sexist insult might be discounted as bad behavior, however, when women attempting to engage in political discourse face a constant stream of harassment and abuse, this becomes violence.”<sup>16</sup> Hence excluded from gendered disinformation are sincere individual expressions of gendered political views or critiques of women in politics which are not weaponised to attack women or reinforce a wider campaign to do so.

We believe this is crucial to distinguish, as otherwise legitimate criticism of gendered disinformation may be rebutted as simply trying to ‘silence distasteful opinions’, and to identify criticism of women or women in public life *per se* as gendered disinformation would be to caricature all women in public life as beyond any legitimate or democratic critique. This line may not always be easy to draw, but the manner of a criticism being expressed, the intended and actual effect, the presence of covert political, economic or social motives, and the wider narratives it invokes or reinforces which all count towards identifying something as gendered disinformation specifically.

This online behaviour, as with online violence against women in politics, operates on several levels to achieve political outcomes: targeting individual women to harm them or drive them out of public life, while also sending a message that women in general should not be involved in politics: that it is too risky, too dangerous, too unsuitable, too detrimental to a state.<sup>17</sup> This message is intended for other women who are in politics or considering it; but also for the demos at large. By playing on existing stereotypes and tropes about women, underlying beliefs that women simply are not good enough for politics are reinforced, with the ultimate aim that women self-censor and fail to garner support from voters or citizens at large.<sup>18</sup>

12. Amnesty International. Toxic Twitter - Women’s Experiences of Violence and Abuse on Twitter, Chapter 3. 2018. Available at <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-3/> [accessed 15 May 20]

13. Wardle, C. and Derakshan, H. Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe, 2017. Available at <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c> [accessed 03 May 2020]

14. Amnesty International. Toxic Twitter. Amnesty International, 2018. Available at <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/> [accessed 03 May 2020]

15. Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 2020]

16. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016, p. 14. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

17. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

18. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

## THE IMPACT OF GENDERED DISINFORMATION: FROM THE PERSONAL TO THE POLITICAL

*'Online violence against politically-active women is one form of the global problem of violence against women in politics (VAW-P), which encompasses all forms of aggression, coercion, and intimidation of politically-active women simply because they are women.'*<sup>19</sup>

NDI, Tweets that Chill

Gendered online violence can have significant political effects, from silencing women in public discourse to reducing women's participation in politics and public life, to seemingly justifying the political persecution of women. NDI's 2019 research in Kenya, Indonesia and Colombia found that Online VAW-P had a 'chilling effect' on politically-active women's social media engagement, driving them to reduce the frequency of their online engagement in political discourse or driving them away from social media platforms entirely.<sup>120</sup> 76% of women polled by Amnesty who had experienced online abuse or harassment reported changing the way they use social media platforms as a result; with 32% stopping posting certain opinions as a result. And a survey by Media Matters for Democracy found that 9 out of 10 women journalists who responded said that online violence affects their professional choices, and 8 in 10 said it led to self-censorship.<sup>21</sup> Women are not only discouraged from running for office, engaging online or speaking up - attacks aimed to discredit them have been used to garner support for critics of state leaders being arrested, put on trial and even imprisoned.<sup>22,23,24</sup>

The personal psychological impact is also significant. Research by Amnesty found that over

*'Gendered online violence can have significant political effects, from silencing women in public discourse to reducing women's participation in politics and public life, to seemingly justifying the political persecution of women.'*

half of women polled who had experienced online abuse or harassment suffered afterwards from stress, anxiety or panic attacks (55%) to trouble sleeping (63%) to lower self esteem (61%), and 41% feeling their physical safety was threatened on at least one occasion.<sup>25,26</sup>

Online violence being targeted particularly at women politicians and women in public life is a widely recognised phenomenon, in a context of widespread psychological and physical violence against women in public life.<sup>27</sup> Political abuse occurs widely against politicians of all genders; however, a survey of UK MPs found that personal attacks and violent threats were much more commonly suffered by women.<sup>28</sup> Women of colour in public life in particular are extremely often affected.<sup>29</sup> Disproportionate attacks on women politicians across Europe have been described as seeking to "target, abuse and humiliate"<sup>30</sup>, focusing on women's physical appearance, making threats and using misogynistic language. 41.8% of global women parliamentary survey respondents report sexual or humiliating images having been spread online.<sup>31</sup>

False and manipulated information can be and is used against people of other genders. This too can carry significant risk and harm to an individual.

19. Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 20]
20. Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 20]
21. Kamran, H. Hostile Bytes. Media Matters for Democracy, 2019. Available at <http://www.digitalrightsmonitor.pk/wp-content/uploads/2019/11/Hostile-Bytes.pdf> [accessed 24 August 2020] via edia Matters for Democracy. Media Matters for Democracy expresses solidarity with women journalists and calls for an end to all kinds of online violence. APC, 2020. Available at <https://www.apc.org/en/news/media-matters-democracy-expresses-solidarity-women-journalists-and-calls-end-all-kinds-online> [accessed 24 August 2020]
22. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 20]
23. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 2020]
24. Perraudin, F. and Murphy, S. Alarm over number of female MPs stepping down after abuse. Guardian, 2019. Available at <https://www.theguardian.com/politics/2019/oct/31/alarm-over-number-female-mps-stepping-down-after-abuse> [accessed 15 May 20]
25. Amnesty International. Amnesty reveals alarming impact of online abuse against women. Amnesty International, 2017. Available at <https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/> [accessed 03 May 2020]
26. Amnesty International. Amnesty reveals alarming impact of online abuse against women. Amnesty International, 2017. Available at <https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/> [accessed 03 May 2020]
27. National Democratic Institute. #NotTheCost: Stopping Violence Against Women In Politics. Submission to the UN Special Rapporteur on Violence Against Women, 2018. Available at <https://www.ndi.org/sites/default/files/NDI%20Submission.pdf> [accessed 03 May 20]
28. Akhtar, S. and Morrison, C. The prevalence and impact of online trolling in UK Members of Parliament. Computers in Human Behaviour 99, 2019. 10.1016/j.chb.2019.05.015 in Shahani, A. Trolled Online, Women In Politics Fight To Hold Big Tech Accountable In The U.K. NPR, 2019. Available at [https://www.npr.org/2019/08/19/751794126/trolled-online-women-in-politics-fight-to-hold-big-tech-accountable-in-the-u-k?utm\\_source=twitter.com&utm\\_medium=social&utm\\_term=nprnews&utm\\_campaign=npr](https://www.npr.org/2019/08/19/751794126/trolled-online-women-in-politics-fight-to-hold-big-tech-accountable-in-the-u-k?utm_source=twitter.com&utm_medium=social&utm_term=nprnews&utm_campaign=npr) [accessed 03 May 20]
29. Amnesty International UK. Black and Asian women MPs abused more online. 2017. Available at <https://www.amnesty.org.uk/online-violence-women-mps> [accessed 15 May 2020]
30. Guerin, C. in Spring, M. and Webster, L. A web of abuse: How the far right disproportionately targets female politicians, BBC, 2019. Available at [https://www.bbc.co.uk/news/amp/blogs-trending-48871400?\\_twitter\\_impression=true](https://www.bbc.co.uk/news/amp/blogs-trending-48871400?_twitter_impression=true) [accessed 03 May 2020]
31. Di Meo, L. Gendered Disinformation, Fake News, and Women in Politics. Council on Foreign Relations, 2019. Available at <https://www.cfr.org/blog/gendered-disinformation-fake-news-and-women-politics> [accessed 03 May 2020]



The risks of gendered disinformation run beyond disrupting people's political life to their personal liberty and safety. Men attacked by disinformation, which may have gendered elements, for instance by being accused of being gay in some countries, can be put at great risk.<sup>32</sup>

However, we focus on gendered disinformation which weaponises gendered narratives with the purpose being to "control, limit or prevent women's full and equal political participation".<sup>33</sup> The individual target of such campaigns is disproportionately (though not necessarily) a woman, as with political online violence against women, which seeks not only to target an individual but to make a general point to other observers, women and non-women: that women should not be involved in political life.<sup>34</sup>

## IDENTIFYING STATE-ALIGNED GENDERED DISINFORMATION

We focus in this report on such politically motivated gendered disinformation: and specifically state-aligned gendered disinformation (SAGD): whereby actors who are part of a state, or whose behaviour or interests align with those of a state, engage in gendered disinformation to promote political outcomes.

Such gendered disinformation specifically has been identified and can be seen in political situations across the world: especially that which "mixes old ingrained sexist attitudes with the anonymity and reach of social media in an effort to destroy women's reputations and push them out of public life".<sup>35</sup> Often this manifests as a state or state-aligned actors using gendered disinformation to advance their own political goals, often to undermine their critics and bolster support within the state, rather than reach an international audience.<sup>36</sup> These campaigns take multiple forms, but with common gendered elements: from

'State-aligned gendered disinformation (SAGD): whereby actors who are part of a state, or whose behaviour or interests align with those of a state, engage in gendered disinformation to promote political outcomes.'

doctored images claiming to be naked photos of women politicians who have spoken out against war to fake photos of women journalists which purport to show they have had plastic surgery.<sup>37,38</sup> Other examples include screenshots purporting to be from women politicians' sex tapes, harassment campaigns by troll armies, and misleading or fabricated content used against prominent women of colour who criticise the state leader.<sup>39,40,41,42</sup>

Identifying state-based gendered disinformation operations - that is, deliberate disinformation campaigns, coordinated by state actors - faces many challenges, in particular, demonstrating coordination and inauthenticity of content, rather than, for instance, many users sharing a similar message which they genuinely believe which so happens to align with the state's interests. Gendered disinformation can, however, be deliberate, without being coordinated. We use the broad term 'state-aligned' to include disinformation created and shared not only by those who may have been directly coordinated by the government, but also by informal networks of actors who use gendered disinformation to try to shield the state from critique or democratic threat, in particular by women in public life.

Challenges still remain of how to identify covert motives in sharing false information or critique, and how to distinguish specifically state-aligned gendered disinformation from more general online violence against women in public life. In this

32. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 20]

33. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

34. National Democratic Institute. #NotTheCost - A Call to Action. NDI, 2016. Available at <https://www.ndi.org/sites/default/files/1%20%23NotTheCost%20-%20Call%20to%20Action.pdf> [accessed July 2020]

35. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

36. Thanks to participants of the roundtable for raising this point. August, 2020.

37. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

38. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 2020]

39. Jancowicz, N. How Disinformation Became a New Threat to Women. Coda, 2017. Available at <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/> [accessed 03 May 2020]

40. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 2020]

41. Posetti, J. Fighting back against prolific online harassment in the Philippines. The Conversation, 2017. Available at <https://theconversation.com/fighting-back-against-prolific-online-harassment-in-the-philippines-80271> [accessed 03 May 2020]

42. Holmes, J. A Cop's Threat Against Alexandria Ocasio-Cortez Shows Propaganda Is Dangerous. Esquire, 2019. Available at <https://www.esquire.com/news-politics/a28479868/alexandria-ocasio-cortez-police-officer-threat-misinformation/> [accessed 03 May 2020]

report, we have approached these challenges in several ways; looking at content targeting people in public life who are likely targets of gendered disinformation and identifying discussion of known instances of disinformation (based on our background review, input from in-country experts and initial examination of the data); looking at politically influential users, many of whom work for the government or are aligned with the state; and looking at their friends and followers to examine the information networks within which these users are operating, as well as identifying commonalities between tweets shared.

Not all of the examples of gendered disinformation we describe in this report come from state-aligned actors (though that is our primary focus): some are even from those explicitly aligned against the state. We discuss the makeup of different groups sharing gendered disinformation in *The Players*. This allows us to present a fuller picture of how online discourse about women is weaponised by state-aligned actors, and examine the overlaps and interactions between state-aligned and other forms of gendered disinformation.

Another aim of this project was to test how effective our methodology was for investigating state-aligned disinformation of a particular type, and whether this method could be refined and widely implemented around areas, topics or events of particular interest.

We do not thus present a conclusive analysis of state-driven gendered disinformation campaigns - we do, however, offer a snapshot of the kind of gendered online discourse, driven by state-aligned actors, that is targeting women in public life.

# CHAPTER 2

# BACKGROUND

This study looks at how state-aligned gendered disinformation occurs in an online ecosystem in two countries, primarily through the examination of Twitter data. We examined tweets sent by users, many of whom were aligned with or influential within the state through their work, connections or social position, and who were identified by in-country experts as particularly influential in online political discourse. We then looked in particular at tweets which used language associated with gendered disinformation and which named likely targets of gendered disinformation, in order to examine how gendered disinformation manifested in online conversations.

## METHOD OUTLINE

Working with the National Democratic Institute, we identified two 'trial' countries to investigate: Poland and the Philippines. These countries were selected based on: the prevalence of disinformation and/or gendered abuse; political attitudes towards gendered issues; levels of online political discourse across different platforms; where there were strong connections with in-country partners to guide the analysis; and language capabilities of the research team. We analysed tweets in Polish and, from the Philippines, in English. All Twitter data was collected through the platform's public API.

Data was collected and analysed as follows:

1. A list of users who heavily influence online political discourse in each country was assembled, and tweets sent by these accounts and a sample of users following them was collected.
2. Tweets were filtered to those containing terms within a lexicon of political and gendered terminology, which was built by in-country experts and refined according to language discovered in the dataset
3. The resulting dataset was filtered to identify tweets containing a name or phrase associated with a likely target of gendered disinformation, based on initial analysis and input from in-country experts.
4. This data was manually analysed to identify which tweets displayed gendered disinformation, and the key themes which arose.
5. Finally, the dataset was used to generate network maps of which users were sharing gendered disinformation, and whether there was evidence of state-alignment in those who were most engaged in this activity.

This project was commenced in August 2019 and completed in August 2020.

In August 2020, a closed roundtable was held by NDI with experts from across the world to discuss the findings and the ramifications for mitigation strategies.

More details can be found in the *Full Methodology*, and recommendations for refining the methodology in *Limitations, Challenges and Methodology Recommendations*.

## COUNTRY CONTEXTS

### Poland

The conservative nationalist Law and Justice Party (PiS), who governed the country since 2015, won the general election in October 2019.<sup>43</sup> Since then, the government and the Prime Minister Mateusz Morawiecki have been criticized by many national and international liberal voices for increasing control over public media and introducing repressive policies, many of which disproportionately impact women. Abortion rights, 'leftist' feminism and LGBT rights have all been

43. BBC. Poland election: Ruling Law and Justice Party win poll. 14 October 2019. Available at <https://www.bbc.co.uk/news/world-europe-50037654> [accessed 15 May 2020]



contested by the governing party.<sup>44,45,46,47</sup> Human Rights Watch have criticised the government's attitudes towards women's rights activists in Poland, on the basis that they were<sup>48</sup> denying funding for women rights charities, advocating for regressive laws and reinforcing traditional gender roles.<sup>49</sup>

Although the use of Twitter in Poland is not high, nevertheless, Twitter has been described by the Stanford Internet Observatory as a place where the terms of political and media debate are set due to its use by political and media figures.<sup>50</sup> Online disinformation and misinformation are widespread in Poland, particularly relating to the rights of LGBT+ people and of immigrants and refugees.<sup>51</sup> Coordinated campaigns to spread far-right content have been identified, and a coordinated online harassment campaign against members of the judiciary who had criticised government policies even led to the resignation of a government deputy minister.<sup>52,53</sup> Communication firms and troll farms engage in astroturfing and attacking political opponents, both of the right and the left: though an Oxford Internet Institute report examining likely fake accounts identified twice as many suspicious right-wing as left-wing accounts.<sup>54</sup>

## The Philippines

Rodrigo Duterte was elected President of the Philippines in 2016. Previously the Mayor of Davao, he is known - and is popular domestically for his hardline approach to crime, particularly drug crime.<sup>55</sup> In the May 2019 elections, his position was strengthened as his allies took control of the senate.<sup>56,57</sup> Human rights in the Philippines is an issue of international concern.<sup>58</sup> The police and security forces, as well as unidentified gunmen, have been involved in thousands of extra-judicial killings relating to the 'war on drugs'.<sup>59,60</sup> Critics of Duterte and his 'drug war', including politicians, journalists and activists, have faced trolling, harassment, attacks, arrest, lawsuits, and detention.<sup>61</sup> Public denunciation is seen as "part of the Duterte style of rule"<sup>62</sup>. Duterte is known for his 'strongman' approach - as well as for his offensive and misogynistic behaviour, including making sexist jokes, making jokes about rape, and attempting to kiss women while on official visits.<sup>63</sup>

The weaponisation of social media by the government and its supporters is well documented<sup>64</sup> - for instance, through "patriotic trolling", where genuine users and paid employees

44. Shukla, S and Klosok, A. Poland debates abortion bill amid coronavirus lockdown. CNN News, 16 April 2020. Available at: <https://edition.cnn.com/2020/04/15/world/poland-abortion-law-intl/index.html> [accessed 15 May 2020]
45. Zbytyniewska, K. Polish activist: Nothing is right about women's rights in Poland. Euroactiv, 2018. Available at <https://wzbytyniewska.com>
46. Zbytyniewska, K. Polish activist: Nothing is right about women's rights in Poland. Euroactiv, 2018. Available at <https://www.euractiv.com/section/freedom-of-thought/news/polish-activist-nothing-is-right-about-womens-rights-in-poland/> [accessed 14 May 2020]
47. Easton, A. Polish election: Leader targets gay rights as threat to society. BBC News, 8 October 2019. Available at <https://www.bbc.co.uk/news/world-europe-49904849> [accessed 14 April 2020]
48. Human Rights Watch. 'The Breath of the Government on My Back'. Attacks on Women's Rights in Poland. Human Rights Watch, 2019 Available at <https://www.hrw.org/report/2019/02/06/breath-government-my-back/attacks-womens-rights-poland> [accessed 12 March 2020]
49. Human Rights Watch. 'The Breath of the Government on My Back'. Attacks on Women's Rights in Poland. Human Rights Watch, 2019 Available at <https://www.hrw.org/report/2019/02/06/breath-government-my-back/attacks-womens-rights-poland> [accessed 12 March 2020]
50. Bush, D. and Kurzynski, M. Poland: Presidential Election 2020 Scene-Setter. Stanford Internet Observatory, 2020. Available at <https://cyber.fsi.stanford.edu/news/poland-scene-setter> [accessed 26 August 2020]
51. Bush, D. and Kurzynski, M. Poland: Presidential Election 2020 Scene-Setter. Stanford Internet Observatory, 2020. Available at <https://cyber.fsi.stanford.edu/news/poland-scene-setter> [accessed 26 August 2020]
52. Bush, D. and Kurzynski, M. Poland: Presidential Election 2020 Scene-Setter. Stanford Internet Observatory, 2020. Available at <https://cyber.fsi.stanford.edu/news/poland-scene-setter> [accessed 26 August 2020]
53. Gałczyńska, M. Sledztwo Onetu. Farma trolli w Ministerstwie Sprawiedliwości, czyli "za czynienie dobra nie wsadzamy". Onet, 2019. Available at <https://wiadomosci.onet.pl/tylko-w-onecie/sledztwo-onetu-farma-trolli-w-ministerstwie-sprawiedliwosci-czyli-za-czynienie-dobra/j6hwp7f> [accessed 26 August 2020]
54. Jancowicz, N. The Growing Threat of Domestic Disinformation in Poland. Center for a New American Security, 2020. Available at <https://www.cnas.org/publications/commentary/the-growing-threat-of-domestic-disinformation-in-poland> [accessed 26 August 2020]
55. BBC News. Profile: Duterte the controversial 'strongman' of the Philippines. BBC News, 22 May 2019. Available at <https://www.bbc.co.uk/news/world-36659258> [accessed 12 March 2020]
56. BBC News. 'Philippine mid-terms: Duterte consolidates grip on power.' BBC News, 22 May 2019. Available at <https://www.bbc.co.uk/news/world-asia-48277482>. [accessed 12 March 2020]
57. Esmaguél II, P. 'Duterte allies named dominant majority, minority parties'. Rappler, 10 May 2019. Available at <https://www.rappler.com/nation/politics/elections/2019/230157-duterte-allies-named-dominant-majority-minority-parties> [accessed 12 March 2020]
58. Human Rights Watch. Philippines: UN Takes Critical Step Toward Accountability, 2019. Available at <https://www.hrw.org/news/2019/07/11/philippines-un-takes-critical-step-toward-accountability> [accessed 15 May 2020]
59. Human Rights Watch. Philippines Events 2018. Human Rights Watch, 2019. Available at <https://www.hrw.org/world-report/2019/country-chapters/philippines> [accessed 12 March 2020]
60. Timberman, D. Philippine Politics Under Duterte: A Midterm Assessment. Carnegie Endowment for International Peace, 2019. Available at <https://carnegieendowment.org/2019/01/10/philippine-politics-under-duterte-midterm-assessment-pub-78091> [accessed 12 March 2020]
61. Human Rights Watch. Philippines Events 2018. Human Rights Watch, 2019. Available at <https://www.hrw.org/world-report/2019/country-chapters/philippines> [accessed 12 March 2020]
62. Timberman, D. Philippine Politics Under Duterte: A Midterm Assessment. Carnegie Endowment for International Peace, 2019. Available at <https://carnegieendowment.org/2019/01/10/philippine-politics-under-duterte-midterm-assessment-pub-78091> [accessed 12 March 2020]
63. Occenola, P. 'Fake News', Real Women: Disinformation gone Macho'. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 10 March 2020]
64. Timberman, D. Philippine Politics Under Duterte: A Midterm Assessment. Carnegie Endowment for International Peace, 2019. Available at <https://carnegieendowment.org/2019/01/10/philippine-politics-under-duterte-midterm-assessment-pub-78091> [accessed 12 March 2020]

boost government narratives online, across multiple platforms.<sup>65</sup> The news organisation Rappler has collected a database of 12 million accounts which post or share disinformation or pro-Duterte material.<sup>66</sup> The structures of “networked disinformation” in the Philippines are more diverse and in some cases informal than in other countries, ranging from professional strategists to online influencers to fake account operators to grassroots volunteers, including a lot of freelancers.<sup>67</sup> We saw this reflected in our examination of user profiles (see *The Players*).

*This report contains quotes of hate speech and harmful speech, including descriptions of sexual and gender-based violence.*

*Where a quote from Twitter could be used to identify an individual user, the quote has been modified to preserve the sense but not the exact words used.*

65. Etter, L. 'What Happens When the Government Uses Facebook as a Weapon?' Bloomberg News, 2017. Available at <https://www.bloomberg.com/news/features/2017-12-07/how-rodriigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook> [accessed 12 March 2020]

66. Etter, L. 'What Happens When the Government Uses Facebook as a Weapon?' Bloomberg News, 2017. Available at <https://www.bloomberg.com/news/features/2017-12-07/how-rodriigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook> [accessed 12 March 2020]

67. Corpus Ong, J. and Cabanes, J.V. Architects of Networked Disinformation. University of Leeds, 2018. Available at <http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf> [accessed 12 March 2020]

# CHAPTER 3

## THE PLAYERS

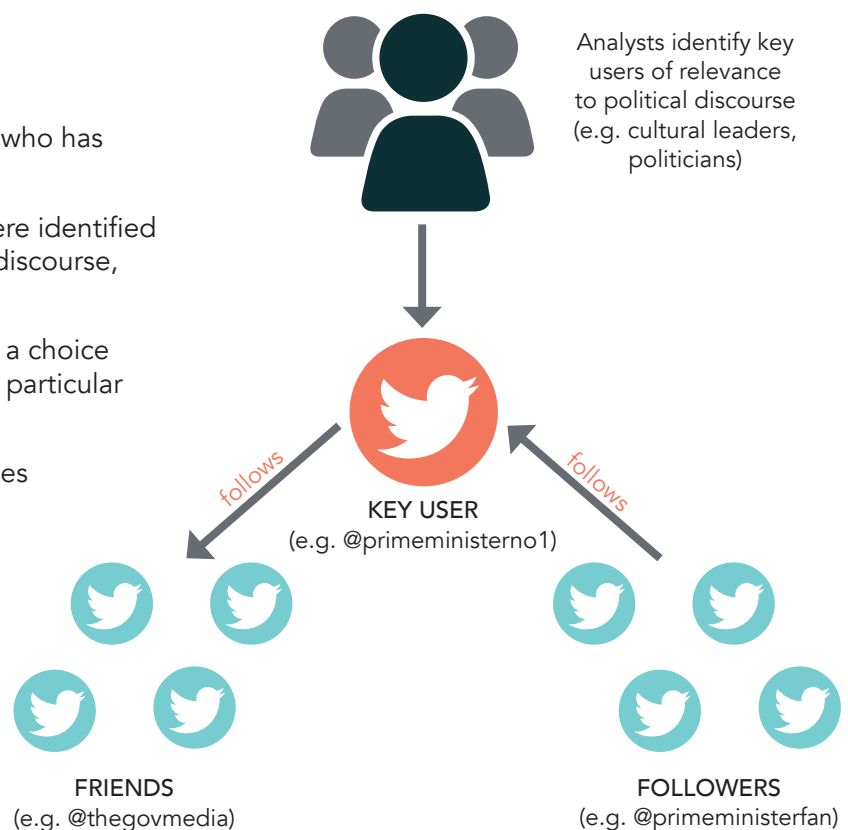
'Social feeds are not a neutral ordering of information.'

We first present a topline analysis of the networks of people online whose online activity we examined to identify state-aligned gendered disinformation (SAGD) in the Philippines and Poland. The majority of gendered disinformation we examined in our research was state-aligned in content - that is, aligned with state interests, attacking critics of the state and so forth (see more in *The Rules Of The Game*). Through this network analysis, we also demonstrate that there is credible evidence of the users who are creating or sharing this gendered disinformation being state-aligned, giving more weight to its identification as SAGD specifically.

To select whose online activity we would focus on, advised by the in-country experts in the Philippines and Poland, we located prominent political and cultural leaders, most of whom were aligned with the state in some way either through their official position or through their alignment of opinions. These were our 'key users'. We collected tweets from them, from their friends and from a random 10% sample of their followers. These tweets produced the dataset which we used for our qualitative analysis (*The Rules Of The Game*).

### Key terms:

- **User:** An individual or organisation who has a Twitter account
- **Key user:** Those individuals who were identified as of particular interest to political discourse, whose networks were analysed
- **Following:** To 'follow' an account is a choice a user makes to see content from a particular other user.
- **Friends:** The accounts a user chooses to follow on Twitter
- **Followers:** The accounts that choose to follow a user on Twitter



One relationship which has a crucial effect on the visibility and spread of disinformation online is that of followership. On Twitter, the accounts a user chooses to follow - their 'friends', in Twitter's terminology - determine the content they are shown on their feed.<sup>68</sup>

Social feeds are not a neutral ordering of information. Rather than displaying, for example, a chronological ordering of tweets sent by a user's friends, Twitter presents content ordered by 'relevance.' Like all social platforms, the algorithms used to determine relevance for tweets are constantly changed and experimented with, but a 2017 blog post by the company explains that one aspect of a tweet which determines its priority on your feed is your relationship to a tweet's author, and your "strength of connection to them".<sup>69</sup>

By following a user, then, you affect the content which you're shown in two ways. In a direct sense, this user's tweets will now appear in your feed, but you are also strengthening the connections between your account and the content your new friend interacts with, meaning this content is more likely to be judged 'relevant' to you.

One tool which can help study these relationships at scale is network analysis. This involves building a graph describing the links between them, and can help gain a high-level understanding of how people are connected online at scale, alongside any distinct collections of people within a larger group. It can also help us understand how SAGD might spread across social networks, and those groups who might be particularly likely to share it.

Our key findings were:

- Gendered disinformation occurring in these countries is often (though not exclusively) created or shared by those who have alignment in some way with the state.
- In both cases, gendered disinformation was clustered particularly strongly in certain groups.

'Gendered disinformation occurring in these countries is often (though not exclusively) created or shared by those who have alignment in some way with the state.'

- In contrast to disinformation, in the Philippines, counterspeech was more widely spread across a network.
- Counterspeech and gendered disinformation are often shared by the same groups of people, though this speech often takes different, politically-inflected forms in different groups.
- In both cases, well-connected networks of relatively low-follower key accounts were found to be responsible for the largest quantities of gendered disinformation, rather than followers of popular political accounts - even when these latter had shared gendered disinformation themselves.
- Gendered disinformation does not always take place within an echo chamber. While the Philippines network shows a group of well connected users sharing gendered disinformation, those sharing this information in the Polish example were well connected to many key accounts, outside of those sharing gendered disinformation.

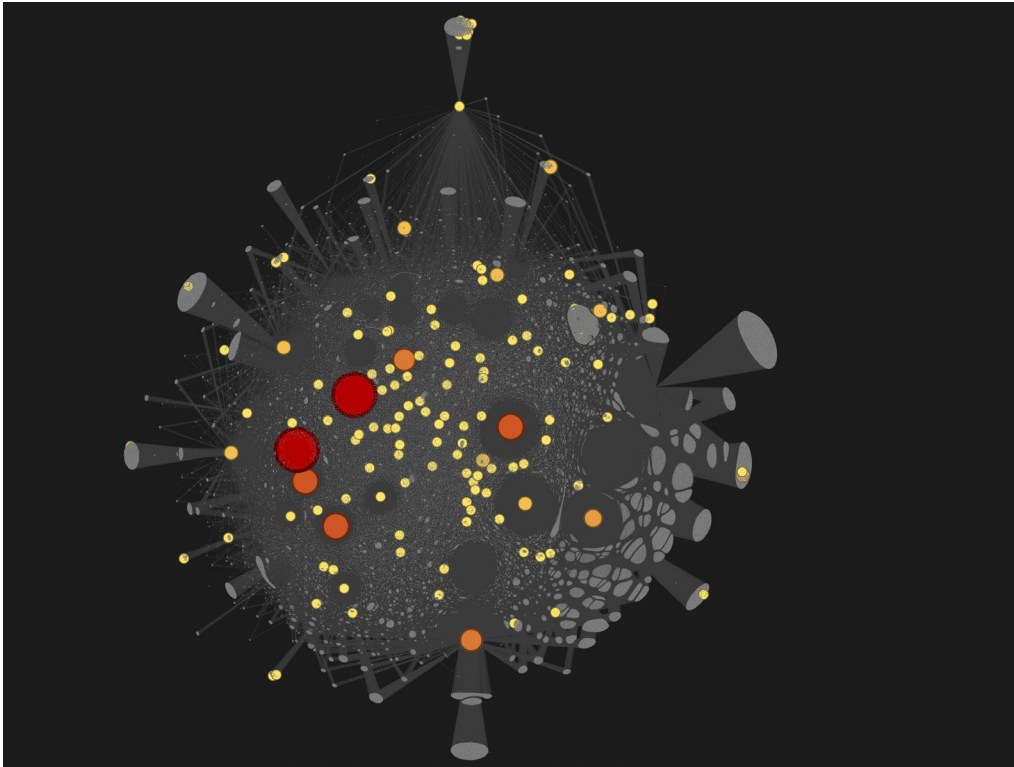
The graphs below map these users out, connecting key accounts (shown in colour below) to their followers and friends (shown in grey, as tiny points) by thin grey lines. These lines are drawn to be as short as possible, so that each user ends up close to those they follow.<sup>70</sup> Key users are represented by 'nodes' or circles. Their followers and friends are represented by tiny grey dots which cluster together in ovoid clumps. These are two of the key graphs we produced: a full discussion of these findings and further findings can be found in the *Appendix: Full Network Analysis*.

68. This may seem odd at face value - Twitter does not use the word 'friends' much on the platform - but it's regularly employed in the platform's technical documentation. See for example the API page for getting statuses of friends here: Developer. Follow, Search and get users. Twitter. Available at <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-friends-ids> [accessed 15 May 20]

69. Koumchatzky, N and Andryeyev, A. Using deep learning at scale in Twitter's timelines. 2017. Available at [https://blog.twitter.com/engineering/en\\_us/topics/insights/2017/using-deep-learning-at-scale-in-twiters-timelines.html](https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twiters-timelines.html) [accessed 15 May 20]

70. Note that we're not drawing lines for connections between two grey dots - we only know which key accounts a user follows, so non-key accounts following each other don't show up here.

## POLAND



**FIGURE II.**  
POLAND - USERS  
COLOURED  
AND SIZED BY  
DISINFORMATION

Different groups of users are well-connected on Twitter: users follow multiple key users, which accounts for the relatively coherent circular shape of the map. This means that messages which are widely shared by one group of followers of key users are more likely to be passed on amongst the circle, raising the possibility that SAGD - as well as counterspeech - could more easily be spread to more users.

Though there are differences (shown in different colours) in how much gendered disinformation different users share, it is shared at a low level amongst most of the network. This map may suggest that the majority of followers of prominent political accounts are less likely to author or reshare disinformation than those who follow multiple smaller accounts picked up in our sample.

We found a higher concentration of users sharing gendered disinformation in certain clusters. A group, which was identified as containing people relevant to political and academic life and the media, accounts for 22% of total users, but it contains 37% of 'relevant' users - those who shared content identified as gendered disinformation. In contrast, another cluster, formed primarily around

political accounts, contains 29% of the total user base, but only 17% of relevant users - under half as many. These relevant users tend to be positioned towards the centre of the graph, and by implication connected to a higher number of key accounts. A third cluster of users sits between the two on the graph, both in terms of its position on the graph, and its saturation of relevant accounts.

Rather than focusing on high-profile, mainstream political accounts known to share gendered disinformation, attention should be paid to smaller, denser networks of users. While each of the 'ringleader' accounts may have a relatively small followership, the above suggests that their audience is more likely to follow other users who have shared gendered disinformation, and are more likely to have shared gendered disinformation themselves.

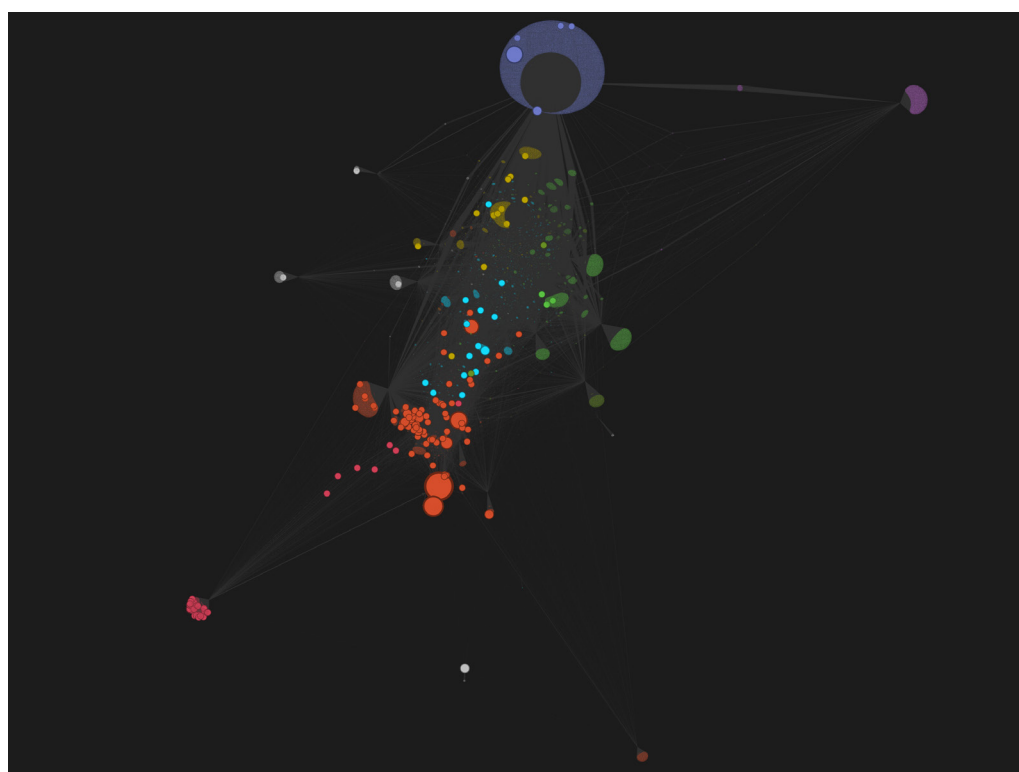
### User profiles

Of the original seed users, those who were identified as sharing messaging relevant to gendered disinformation included: 16 unique users (some of whom posted multiple times). Seven of these were involved in political life. Five others

were political operatives, several of whom were also prominent in state-aligned institutions or broader culture. Four were relevant either to politics, society, or institutions. Only one was verified on Twitter. Users reported they were in seven unique locations (or no location), all of which were in Poland. Their friends count ranged from 142 to 5,918 and they had between 1,712 and 187,588 followers. The average number of friends was 1,141 and the average number of followers was 64,901.

Of the followers and friends sample, those who were identified as sharing gendered disinformation messaging included: 115 unique users, none of whom were verified users, across 28 unique location values. The majority of these were in Poland, but included Germany and the USA. The follower count ranged between 0 and 7,447 and the friends count ranged between 6 and 5,398. The average number of followers was 309, and the average number of friends was 461.

## THE PHILIPPINES



**FIGURE VI.**  
PHILIPPINES -  
TWITTER USERS  
GROUPED BY  
MODULARITY

This map shows that there are groups of key users who are spaced distantly from each other, with users in closely knit groups much less likely to follow relevant users from other groups. Users sharing gendered disinformation tend to be concentrated in one area of the network - a group separated from other key users towards the lower right of the graph (larger orange nodes). This may represent an echo chamber - users who are more likely to follow others in the group than other key users elsewhere on the graph, and who are likely to see, send and share high quantities of gendered disinformation.

The 'gold' accounts near the top of the map are related to politicians who are known to speak out against gendered harassment, gendered disinformation and abuse, as well as supportive human rights activists. Lying right under this cluster (suggesting they share followers) is a 'light blue' group, composed mostly of politicians, though also including media accounts and personalities. Some of these are more explicitly aligned with the government, while some are critical of the government.



Below this, the tightly clustered 'orange' accounts are commonly associated with cultural influencers and bloggers, mostly pro-Duterte and/or critical of the opposition. They show the 'revolving door' that exists between positions of political power and government authority, and those who are influential in online media and cultural discourse: some accounts are of those who have held both political and media roles at different times. Others are organisations, private individuals who comment on political affairs, and online political communities.

These numbers show a clear difference between groups. The orange, pro-Duterte group makes up a mere 13% of users on the graph, but account for 61% of users sharing gendered disinformation - though this cluster also accounts for 23% of users sharing counterspeech. In contrast, the Gold cluster, primary account following politicians speaking out against gendered harassment, accounts for 7% of the population and of gendered disinformation - but 21% of counterspeech.

In the 'orange' group, the group of most interest in analysing SAGD specifically, hashtags shared in tweets which used a relevant term (from the lexicon) often focused on undermining women politicians, and also on supporting the government. "#leniresign" was used 10 times, "#lenifakenews" six times, "#fakevprobredo" four times, and "#protectourpresident" (in conjunction with other hashtags) 10 times. The conjunction of "#lenipowergrabber" and "#protectprrd [Duterte]" five times in the dataset implies a connection being made between undermining women in politics in the ways set out above, with the 'protection' of the state. Interestingly, this group also used #takebackthetech six times - a hashtag more commonly associated with counterspeech. In the 'gold' group of accounts, hashtags focused on speaking up against violence against women ("#metoo", six times; "#violence against women", four times; and "#endrapecultureph", 12 times), as well as critiquing the state.

The hashtag usage of these clusters alone support our identification of two different groups of particular interest: those who are broadly in support of or aligned with the state, who also join in with online conversations undermining prominent women in politics; and those who are further from the state and engage more clearly in

counterspeech. This also appeared as a pattern in the word frequency of tweets shared by users in these groups, though the groups were not absolutely separate. Terms appeared across all three groups, but often terms more associated with SAGD (e.g. boba, presstitutes, saba) appeared more often in the 'orange' group, and terms more associated with counterspeech (misogyny, sexual violence) appear more often in the 'gold' group, with the 'blue' group as a middle ground.<sup>71</sup>

### User profiles

Of the original seed users, those who were identified as sharing gendered disinformation messaging included: eight unique users (some of whom posted multiple times), five of whom had been categorised as involved in culture - cultural leaders, bloggers - reflecting the in-country team steer that it was these 'informal', popular, often state-aligned sources who were the most likely to be spreading gendered disinformation. Two were more directly related to state operations. Another represented a pro-Duterte online community. The number of friends these users had (people they followed) varied between 26 and 829, and the number of followers varied between 142 to 78,443. The average number of friends was 337, and the average number of followers was 17,229. Hence while not the largest presences on Twitter, some have significant networks.

3 users shared news relevant to gendered disinformation including a target of it.

Of the followers and friends sample who we identified as sharing gendered disinformation, there were 216 unique users. None were verified; their friends ranged between 1 and 4,859, and their followers between 0 to 17,614. They were reportedly from 89 different locations (note that these are set by the user themselves so may not necessarily be accurate - indeed, some were jokes) - many were from the Philippines, but there were also users who said they were located in the US, Europe and Asia. The average number of friends was 303, and the average number of followers was 773.

Of the original users who engaged in counterspeech (understood in the broadest sense, of speaking against violence against women), there were 16 unique users. These included some of the original users who shared gendered disinformation,

71. "Bitch" appearing on its own was found only in the orange group, but phrases which included the term e.g. "yes bitch" were found in other groups.

showing that there are not always 'clear sides' that people fall on. There was more political representation in this set - there were again several 'cultural leaders' or bloggers (seven) but also three politicians as well as the two state affiliated actors identified in the first set. The friends range in this group was between 14 and 1,317, and the followers group 72 to 155,447.

What this shows is that gendered disinformation occurring in these countries is often created or shared by those who have alignment with the state.

Given the non-representative initial selection of the accounts we examined, we cannot say that it is always or most often, (and indeed is not exclusively) shared by state-aligned actors. However, this does point to gendered disinformation being one way in which state-aligned actors are engaging online.

How state-aligned gendered disinformation occurs online, however, is extremely context-dependent: not only do different groups of people hold power to shape online political discourse, but the ways in which they are connected with other users of an online platform varies hugely. The range of people who engage in gendered disinformation relating to a particular country is wide, both geographically

and in terms of online reach and influence - from significant social or political influencers to individuals with small followings. This supports an understanding of the methods and operations of gendered disinformation whereby messages shared by powerful influencers, overtly or covertly, spread out to reach a wide audience who in turn co-opt the messaging and make it their own.<sup>72</sup>

Moreover, they highlight the challenge of identifying state-based gendered disinformation: as those who engage in SAGD are often interacting in smaller, less prominent networks - or in account clusters adjacent to but not coextensive with governmental account clusters.

Solutions to state-aligned gendered disinformation, therefore, must take into account where SAGD is occurring in a network. Sometimes it may be clearly clustered in one 'echo chamber' community online; other times it will be spread out across different online communities.

However, there are still commonalities between how SAGD is presented online which may assist in its identification even across different network systems. We present these below.

72. Corpus Ong, J. and Cabanes, J.V. Architects of Networked Disinformation. University of Leeds, 2018. Available at <http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf> [accessed 12 March 2020]



# CHAPTER 4

## THE RULES OF THE GAME

We have thus considered who is engaging in state-aligned gendered disinformation and counterspeech online. We now turn to *how* they do so.

State-aligned gendered disinformation follows clear patterns. These are not incidental or coincidence. Rather, there are ‘rules’ established as to what an actor should do and how they should behave in order to pursue a political agenda of excluding women from public life. These ‘rules’ may be prescribed in more formal networks of disinformation, or observed and copied or mutually defined by informally connected actors. This was found in an examination by the NDI of online violence against women in politics in Indonesia: where “when a user experiences violence on the platform and other users quickly join in, either through amplifying the original violence through ‘re-tweets’ or compositions of their own”.<sup>73</sup>

Those engaging in disinformation are not only trying to influence people’s thoughts but the very parameters and contours of discussion - networks of disinformation in the Philippines have been compared to an ad agency, trying to control how people view and talk about a brand through the spread of particular messaging.<sup>74</sup> Through engaging in online political discourse in set ways, whether that is the use of certain words, hashtags, tropes or targeting particular people and groups, the rules of online political discourse itself are re-shaped, into a language game which excludes women as full, active and respected members.

‘State-aligned gendered disinformation follows clear patterns. These are not incidental or coincidence.’

It is through this lens that we present six ‘rules’: these are patterns we consistently observed in this research in how women in public life were attacked and spoken about, and gendered narratives which were being manipulated by state-aligned actors and their networks. We focused on Poland and the Philippines specifically, but these rules have echoes globally in the kinds of abuse women are subjected to and the narratives which authoritarian leaders aim to promote to consolidate their own power and political agendas. Their aim is to undermine women in public life, to bolster their critics, and to demonstrate the power that can be brought to bear on women who step out of line. Taken together, they call for women’s role in politics, democracy and society to be reduced or removed.

Identifying how far these tactics are deliberate (i.e. deliberate instructions to disinformation agents) and how far they have organically developed through gendered disinformation-aligned conversations online is a challenge, but describing them demonstrates likely patterns that conversations which attack or denigrate women are likely to take.

We primarily examine state-aligned content - that is, which was shared by a known state-aligned actor, or which aligns with the state’s political goals or

73. Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019, p. 17. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 2020]

74. Ong, J.C. and Cabanes, J.V. Architects of Networked Disinformation. University of Leeds, 2018. Available at <http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf> [accessed 12 March 2020]

attacks critics of the state, or which uses a trope already well documented as being part of a state disinformation campaign.<sup>75</sup> However, not all of the actors who engaged in gendered disinformation, whose tweets we discuss below, are aligned with the state: some are even explicitly aligned against the state. We include these as they demonstrate to an extent the permeability of online discourse: how tropes used by state-aligned actors reflect and reinforce those used elsewhere in online conversations rather than create them anew, and how those perpetuating and countering violence against women may sometimes be the same people in different contexts. This highlights a key challenge in combating gendered disinformation: that the boundaries are not always precise.

By identifying messages shared by key influencers of online political discourse, which weaponise gendered language and narratives against known targets of gendered disinformation, we are able to characterise how the conversation about women in public life is being shaped.

#### **RULE 1**

#### **Convince others that women are devious: they are not fit for politics**

A central narrative in gendered disinformation is that women are not fit for public or private political lives, and as such, should be excluded. Most frequently, this language surrounds their character: as presented through gendered disinformation, women are devious, unreliable and fake.

“Fake news” has been weaponised. Originally coined to combat disinformation, it is now part of the populist playbook, and is a widely-used tool in attacking women. Women journalists are accused of being “fake news”. Claims that they are funded through dubious or corrupt means are common. Women politicians are said to lie about their motives and about their political principles, deliberately playing on cultural stereotypes that women are fickle and untrustworthy. It is also a highly powerful tool - calling someone a liar then works to invalidate any past and future claims they can make in the eyes of persuaded observers, without having to refute each claim individually.

Treason and disloyalty emerged as a major theme in the Polish gendered disinformation data. Accusations of actions against “Church and state” were frequently levelled at women politicians in

the country on Twitter. Women were accused of harbouring loyalties to Russia, to feminist principles, to leftism and communism and to powers hostile to the Polish state. Anti-Semitic conspiracy thinking was also represented. *Zdraja* (traitor) and *zdrada* (treason) were common insults. Janusz Korwin-Mikke, a far-Right politician in Poland famously described international women’s day as a Communist invention. Elsewhere, Aleksandra Dulkiewicz, the mayor of Gdansk and a common target of online attacks, has reported being targeted by disinformation alluding to anti-Polish intentions, such as a faked photo of her wearing a Nazi helmet and the flag of the Third Reich.<sup>76</sup>

Women in public life in the Philippines are also frequently portrayed as liars, devious and criminal. Journalists who criticise Duterte, women and especially those who work at Rappler, a woman-led organisation, are described as “presstitutes”, “fake news”, “fake reporter” and “paid hacks” - implying they are playing at journalism and only pretending to be reporting serious issues for the benefit of some mysterious funder (“yellow funded presstitutes”). Some have gone further than devious but portrayed them as “an evil genius” for unleashing “hell” on others. Leni Robredo, vice-president of the Philippines, who was fired from heading up the government’s anti-drug agency after criticizing extrajudicial killings, is frequently attacked using a variety of hashtags and phrases related to “fake VP” and “#FakeVP” and has been called a “cheater” and “liar”.<sup>77</sup> Charges of being fake or lying are often used in conjunction with “bitch”, emphasising that it is a nasty woman who is dishonest: “lying bitch” “fraudulent monster bitch” “despicable greedy hypocrite bitch”. Criticism of others includes charges that they only “[pretend] to be relevant”; and when women speak out against their abuse, they are told their claims are “bull” (see Rules Two and Three).

In a similar vein to the portrayal of women as evil, scheming liars, we also found women being portrayed as being badly motivated - having “ulterior motives”, “motivated by self interest” “not [motivated] by love of country”, working not for the general good or the public interest as they claim, but to pursue their own agenda, which is often as simple as their own glorification “she’s simply a bitch and just wants the attention all the time”. This latter charge was found in both the original seed user dataset and the follower and friends dataset.

75. However, we also found gendered disinformation being used against women who were aligned with the state.

76. TVN24. Mayor of Gdańsk received offensive fake photos. Prosecutor’s office drops the case. 2020. Available at <https://tvn24.pl/tvn24-news-in-english/mayor-of-gdansk-received-offensive-fake-photos-prosecutors-office-drops-the-case-3046045> [accessed 15 May 20]

77. Santos, A. P. “Are you afraid?”: Philippines’ Robredo defiant over drugs war. Al Jazeera, 2019. Available at <https://www.aljazeera.com/news/2019/11/afraid-philippines-robredo-defiant-drugs-war-191126043336589.html> [accessed 30 July 20]

“Traitor” was also levelled against women in the Philippines (“crazy bitch traitor” “surely [she] is a traitor”).

The impact of this is not simply to undermine these women’s standing, to attack their reputations or cause them individual psychological harm. These attacks aim to turn the truth on its head; and the cumulative effect of these repeated attacks on women as being liars and frauds is to paint a picture within which women who are prominent in public life are not to be trusted with such an important role. Not only are they wrong about things - they are deliberately and wickedly scheming to manipulate the wider public - playing into long-held stereotypes about women as witches or succubi. It also serves to protect those who attack and undermine women, since criticism by a woman who is dishonest cannot be legitimate. This echoes a known strategy of political disinformation identified in our report *Warring Songs* - to critique opponents in order to ensure that one’s own political position remains strong.<sup>78</sup> Their very existence as women in public life must be a sham - because women do not belong there.

*nazwanie bohatera zdrajcą to doły.  
nazwanie go zdrajcą przez zdrajcę to  
najgorsze suka (sorry).*

*Calling a hero a traitor is awful. Calling  
a hero a traitor when you yourself are  
the traitor is the worst, bitch (sorry).*

*[She] said: ‘I have no plans to replace  
President Duterte’ - Liar, liar, liar - that lying  
bitch is lying about her ulterior motives.*

## **RULE 2: Denounce women as too stupid for public life**

Through a mix of cited pseudoscience, old-fashioned ‘humour’ and general abuse, the stupidity of women was a regular topic of gendered disinformation in the datasets we examined.

In *Politics*, Aristotle saw women as inferior to men, lacking the authority to govern.<sup>79</sup> Centuries later, populist leaders around the world continue to repeat these tropes, with particular ire reserved for women who challenge a male political hegemony. Many examples can be given from around the

world: Trump lashing out at a woman of colour journalist and insulting the reporter’s intelligence and competence in 2018.<sup>80</sup> The deputy Prime Minister of Italy Matteo Salvini comparing the female president of the lower house of parliament to a sex doll in 2016.<sup>81</sup> And Jair Bolsonaro, President of Brazil, mocking women opposing his views as stupid and “unworthy of rape”.<sup>82</sup> This approach to undermine women in politics and public life was noted by the in-country team in the Philippines, and is borne out in our dataset.

In the Philippines, there was a great deal of discussion about news stories, and notably around political leaders (mostly Duterte) making sexist or misogynistic comments - from rape jokes to insulting world leaders as “sons of whores”. This sparked, not only discussion of the news stories themselves, and analysis of what they meant for Duterte (often that he was still ahead, or was the focus of discussion despite his remarks) - and admittedly a lot of counterspeech denouncing the speech - but also uses of this news in order to further political attacks on women.

During the period researched, a male politician in the Philippines called the woman Vice President of the Philippines, “boba” in a tweet.<sup>83</sup> ‘Boba’ is one commonly used gendered slur, meaning ‘someone (a woman) who is stupid’. In our dataset of most politically influential seed users, one user sent four different tweets in response to this: from making comments that it was not only accurate (‘descriptive’) but that it was legitimate given the “name-calling” of Duterte that his critics engage in, to saying that he had ruined the joke by saying it because it was a “national secret joke”. These tweets received between 22 and 414 favourites (the third most favoured tweet in the set), and 3-89 retweets - a total of 727 favourites and 139 retweets.

These tweets were shared widely among the followers and friends of the influential accounts: 17 tweets (copied or retweeted from the original tweet) included a news story from the Manila Bulletin on the incident alongside the text of the original “boba” comment from among our friends and follower network alone. These tweets did not directly comment in support or critique of the comment, and so intention is hard to discern -

78. Krasodomski-Jones, A. and others. *Warring Songs: Information Operations in the Digital Age*. Demos, 2019. Available at <https://demos.co.uk/project/warring-songs-information-operations-in-the-digital-age/>. [accessed 13 May 20]

79. Aristotle. *The Politics* (Classics). Penguin Classics, 2000.

80. Morin, R. Trump draws rebukes for scolding female reporters of color. *Politico*, 2018. Available at <https://www.politico.com/story/2018/11/09/trump-cnn-white-house-access-980280>. [accessed 13 May 20]

81. Beinart, P. The New Authoritarians Are Waging War on Women. *The Atlantic*, 2019. Available at <https://www.theatlantic.com/magazine/archive/2019/01/authoritarian-sexism-trump-duterte/576382/> [accessed 13 May 20]

82. Philips, T. Stop this disaster: Brazilian women mobilise against ‘misogynist’ far-right Bolsonaro. *The Guardian*, 21 September 2018. Available at <https://www.theguardian.com/world/2018/sep/21/brazilian-women-against-jair-bolsonaro-misogynist-far-right-candidate> [accessed 13 May 20]

83. ABS-CBN News. Locsin calls Robredo ‘boba,’ then says sorry for being ‘disrespectful’. *ABS-CNN News*, 2019. Available at <https://news.abs-cbn.com/news/06/24/19/boba-locsin-blasts-robredo-over-comments-on-diplomatic-passports-cancellation> [accessed 13.05.20]

however, by retweeting the original comment and including the full text of the insult, the message and normalisation of that as an insult towards the Vice President was extended.

*What we've all been saying for longest time, is now even confirmed in the government.*

*Indeed, Leni is boba and has not got a brain.*

*The Vice President is embarrassing  
[translation]*

Although the politician apologised soon after to Robredo, this shows how a comment online by a government official or by a political influencer can trickle down by virtue of having the exposure and political spotlight, bringing to light something which previously was seen as 'hidden'.

It is worth noting that this event sparked criticism and counterspeech online - our dataset included affirmations of Robredo's "beautiful mind" (itself invoking a gendered term) or users chastising or even attacking the original speaker.

This theme - of women being stupid and in particular brainless, came up again and again in our dataset, often in the context of vitriolic language, and was used to reinforce false stories or rumours about women politicians. "Can't even organize her thoughts properly", "incompetent", "brainless bitch[es]", "dead useless brain", "senseless", "boobs for brains", "shrunked brain": these were used mostly against women who were critical of the state, but also state-aligned prominent women. We saw tweets claiming that a politician's rumoured sexual behaviour had led to her making bad decisions - "is this what saba [banana] did to her brain?" asked one user about a senator's policy position. This was attacking Leila de Lima, often referred to with the term 'saba' as this term was one used in de Lima's hearing, as a witness claimed that her affair involved she and her partner feeding each other bananas.

Any misstep that women make is also instantly pounced on and not easily forgiven - or even fabricated where no mistake was made. We found references in the dataset to Robredo not being able to do maths, and calling her "boba" or an "insult" as a result - presumably a reference to when Robredo was speaking about the cost of rice, after which many accused her of making a multiplication error. A prominent journalist was misquoted regarding a statement on rape threats against her:

this misquoting was then also used against her, with users online commenting that maybe she misunderstood or was exaggerating the threat - simultaneously invoking that she had made a stupid mistake while also disregarding the existence of the original abuse by compounding it. Thus, that women are too stupid to understand things is used as a tool to counter allegations of existing abuse and harassment.

And stupidity is not the only attack on women's brains that people make. "Gaga" and "crazy" were also used frequently (e.g. "crazy bitch traitor" "idiot...stupid gaga") - a well-known trope for dismissing women, their beliefs and experiences.

*Still, your Boba VP cannot take over – she is INCOMPETENT*

In 2017, a Polish MEP went on a rant in Parliament over the weakness of women, saying that "Of course women must earn less than men because they are weaker, they are smaller, they are less intelligent".<sup>84</sup> This comment re-emerged during the data collection and was widely shared. The dataset as a whole was littered with accusations of stupidity around decisions made by Polish women politicians, most commonly in response to news articles reporting those decisions.

*@user Idiotyczna dziwka bez szkoły*

*@user Stupid uneducated slut*

These tactics aim to show that women are not thinking correctly - and that therefore what they say should not be believed. The message of these tweets is similar to that of Story One: that women are not to be believed, that they are not fit for office. Whether it is deliberate deceit, physical or mental illness, lack of intelligence or knowledge, women are seen as being unable to be competent or experts.

### **RULE 3: Make women afraid to talk back**

Over the past five years, platforms have made changes aimed at improving the experience of vulnerable users, though targeting of women in coordinated online abuse continues today. Women calling out this harassment, or using tools to reduce it through blocking or muting accounts, was often mentioned in abuse. On top of this, platform decisions seen to be protecting women were also decried.

84. The Telegraph. Sexist Polish MP suspended and fined for saying women should be paid less. The Telegraph, 2017. Available at <https://www.telegraph.co.uk/news/2017/03/14/sexist-polish-mep-suspended-fined-saying-women-should-paid-less/> [accessed 13 May 20]



In the Poland dataset, tweets decried the removal of content by the platform, and on other platforms such as Facebook, as protecting the “sodomite standards” and “feminist perverts” in that online space. Women who blocked their abusers were called out as cowardly and afraid of debate, or that they were hiding from those looking to hold them to account. We noticed a pattern whereby an unblocked user would post about a prominent woman, and users that had already been blocked would join the conversation, most often citing the fact they had been blocked as further evidence to the group that the woman was duplicitous, scared or an opponent of free expression.

*@user zamknij się kurwo - teraz na pewno mnie zablokujesz. tchórz :)*

*@user Shut up whore - you'll block me now of course. coward :)*

In the Philippines, women journalists - particularly those who are critical of the Duterte's government - are the primary targets of disinformation campaigns,<sup>85</sup> dubbed “presstitutes” for their so-called duplicitous and compromised reporting against Duterte.<sup>86</sup> The label “presstitute” was used in our dataset continuously as a method to discredit journalists or the media, sometimes suggesting that the media as an institution had been taken over and compromised by these reporters. Tweets would directly discredit reporters, suggesting that “Maybe they are not presstitutes - maybe they are just inept, stupid people that think they're journalists”. Many, though, would be more clearly directed at discrediting the female journalists for “telling-a-lie to the public” about Duterte or his policies. Some would go further and call for their removal from public life, along with other enemies of the state regime: for example, “All bad things must come to an end! Liberal Parties, Presstitutes, Drug lords, Narco politicians, Criminal” and sometimes with physical call to actions, such as those that would instruct others to kick “the butts of presstitutes”.

That the term is so widely used is not unique to the Philippines, has been traced back to General Vijay Kumar Singh, a politician and former Army general in India,<sup>87</sup> where critics claim it has become a regular part of hate speech against female journalists on social media.<sup>88</sup> In the Philippines, journalists argue that the term can be used by both sides of the information war - criticising pro-Duterte social media users for spreading disinformation themselves - although our dataset suggests it is rarely used as counter speech.<sup>89</sup>

Maria Ressa, the CEO of the online news site Rappler, has received exceptional online abuse for speaking back to power and has regularly been called “presstitute” in our dataset.<sup>90,91</sup> She is now world famous for speaking out against the “propaganda techniques” that troll those who are deemed non-compliant with the status quo in the Philippines with words like presstitutes. We found tweets that would directly discredit her attempts to talk back, for example, calling her a “bitch” that “keeps on saying that she is being oppressed when she isn't” or rejecting her claims that pro-Duterte accounts have been sending her rape threats, calling it “fake news”. There seems to be a feedback loop on Twitter, as the more women speak back against disinformation, the more abuse they receive for their defences.<sup>92</sup>

Frequently also observed in the Philippines data were tweets aimed at internationally famous women, not from the Philippines, who had commented on the political situation in the Philippines - from celebrities to UN Special Rapporteurs. These women were attacked in a similar way to others - called “bitch” “old witch” “idiot” or “gaga”. Their being not from the Philippines was specifically weaponised, as they were portrayed as meddling outsiders who didn't know what they were talking about - “you are not welcome”, “our business not yours”, “do your research b4 you talk”.

85. Combinido, P. When illiberal social media takes over democratic Philippines. New Mandala, 2019. Available at <https://www.newmandala.org/when-illiberal-social-media-takes-over-democratic-philippines/> [accessed 13 May 2020]
86. Sombatpoonsiri, J. Manipulating Civic Space: Cyber Trolling in Thailand and the Philippines. GIGA Focus, Asia, Number 3. 2018. Available at [https://www.giga-hamburg.de/en/system/files/publications/gf\\_asien\\_1803\\_en.pdf](https://www.giga-hamburg.de/en/system/files/publications/gf_asien_1803_en.pdf) [accessed 13 May 2020]
87. Valderama, T.C. Who's the real “presstitute?”. The Manila Times, 2016. Available at <https://www.manilatimes.net/2016/09/25/opinion/columnists/whos-the-real-presstitute/287885/> [accessed 13 May 20]
88. David, S. #Presstitute: The Online War Against Women With An Opinion. The Citizen, 2017. <https://www.thecitizen.in/index.php/en/NewsDetail/index/7/10942/Presstitute-The-Online-War-Against-Women-With-An-Opinion> [accessed 13 May 20]
89. Valderama, T.C. Who's the real “presstitute?”. The Manila Times, 2016. Available at <https://www.manilatimes.net/2016/09/25/opinion/columnists/whos-the-real-presstitute/287885/> [accessed 13 May 20]
90. Etter, L. What Happens When the Government Uses Facebook as a Weapon?. Bloomberg Businessweek, 2017. Available at <https://www.rappler.com/nation/148007-propaganda-war-weaponizing-internet> [accessed 13 May 20]
91. Posetti, J. Fighting back against prolific online harassment in the Philippines. The Conversation, 2017. Available at <https://theconversation.com/fighting-back-against-prolific-online-harassment-in-the-philippines-80271> [accessed 13 May 20]
92. “Pile-ons” where Twitter users jump onto and join in with existing abuse was also found in a study of online violence against women in politics in Indonesia: see Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 20]

There was some criticism also of the 'Bawal Bastos' law against gendered harassment, which echoed this critique of women standing up to harassment as somehow 'hypocritical'.<sup>93</sup>

The anti-bastos law was permission for women to dress bastos?!...Now women have permission to dress like a slut and not to be called a slut...

The purpose of this story is not only to further discredit public women to outside observers, and try to discredit their statements (which in the case of internationally famous women and politicians, may reach a huge audience), but to send a message to the women themselves - if you talk back, we will redouble our efforts against you.

**RULE 4:**  
**Praise women for being sexy, condemn them for being sexual**

Sexualised distortion is common in disinformation campaigns against women - from sharing doctored photos of women politicians naked, to screenshots purporting to be from sex tapes, to accusations of illicit affairs. The power of these tactics comes from social norms which pervade about women and their sexuality - from the 'virgin' archetype that women should remain pure, to the 'whore' - that women are obscenely sexual and will use their sexuality to get what they want. Manipulated or misleading sexual images alleging to be of women in public life and rumours about affairs and pregnancies are frequently circulated - a clear and persistent example of SAGD as sexualised distortion.

The examples below are consistent with sexualised distortion as a tactic seen used against women globally. Although they may appear very 'personal', in that they are talking about (or purport to talk about) women's private lives, they appeal to norms and ideals about how women 'should' behave which can influence voter behaviour<sup>94</sup> : demonstrating how what may appear as simple rumour can be driven by significant political motivation.

These images and associated commentary have even been used to give 'credence' to criminal allegations against women critical of national leadership such as Leila de Lima in the Philippines.<sup>95</sup> Much of the discussion we saw around Leila de Lima online focussed on her alleged affair - with

her often even not referred to by name, but by the term "saba queen". Tweets accused her of giving out sexual favours to her security detail and drivers and she was described as a "whore that satisfy her sexual desires". These kinds of discussion operate on multiple levels: they reinforce entrenched gender norms about how Filipino women 'should' behave, while also leveraging those norms to discredit a critic of the government and bolster the campaign against her.

In our data set, some tweets, which were on the border between counterspeech and gendered disinformation appeared more sympathetic to women targets, but still undermined them on the basis that women were too sexual or emotional: showing that it may not only be a woman's direct opponents who engage in SAGD, but that these narratives may percolate out to others who have more sympathy but still amplify the misleading and politically powerful message.

*Humans make mistakes because of love and sex - de Lima is only one of them.*

*@user De lima is whore that satisfies her sexual desires and is involved with the drug pushers - it was proven that is why she is in jail...*

And it was not simply those targeting women critics of Duterte who engaged in these narratives. Mocha Uson, a prominent Duterte supporter, pop star and blogger,<sup>96</sup> who has worked in several government positions,<sup>97</sup> was frequently the target of this sexualised distortion - even as people criticised her for attacking other women. Tweets expressed the sentiment that she had no business being involved in politics and she should stick to being a "slut", a "mere sexy starlet", "filthy bitch" "cheap whore", "bitch" and engaging in sexual activity. One user said they had more respect for her when she was "just a slut" than her "propaganda" work. A tweet, which appeared several times with different Twitter users mentioned in the tweet each time, asked why anyone would bother with her, a "prostitute" pretending to be intelligent, who was only good at a long list of sex acts. There were many shares of the news that the Mocha Girls, a group of which Mocha Uson was one, would give up "sexy shows" in a show of support for Duterte.

93. Torregosa, H. Hontiveros welcomes signing of IRR for 'Bawal Bastos' law. Manila Bulletin, 2019. Available at <https://news.mb.com.ph/2019/10/28/hontiveros-welcomes-signing-of-irr-for-bawal-bastos-law/> [accessed 13 May 20]

94. See, e.g. Bjarnegård, E. Men's Political Representation. Oxford Research Encyclopedia of Politics, 2018; and Edström, J. Undressing Patriarchy in the Male Order Development Encounter in Engaging Men in Building Gender Equality, ed. Flood, M. and Howson, R., Cambridge Scholars Publishing, 2015, pp. 71-85.

95. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 20]

96. Chen, A. When a populist demagogue takes power. The New Yorker, 2016. Available at <https://www.newyorker.com/magazine/2016/11/21/when-a-populist-demagogue-takes-power> [accessed 13 May 2020]

97. Occeñola, P. Fake News, Real Women: Disinformation Gone Macho. Rappler, 2018. Available at <https://www.rappler.com/newsbreak/in-depth/217563-disinformation-gone-macho> [accessed 03 May 20]

The Polish data focused on the importance of preserving the purity of Polish women. The dominant theme was in the importance of severe punishments for rape, couched in language insisting on the need for strong men to protect women. Tweets in this category ranged from personal anecdotes about beating up people, including likely apocryphal stories of attacking migrants who were harassing a woman, through to explicit calls for more stringent punishments for sex crimes against women. Much of the language was heavily conspiratorial, with a focus on paedophilia within the Church and political opposition and emphasising an association between refugees and migrants and sexual crime.

[LINK] *trzeba chronić polskie kobiety. o zaostrenie kary za gwałt - podpisz: [LINK] #rapefugees*

[LINK] *We must protect Polish women. To increase the penalty for rape - sign here: [LINK] #rapefugees*

#### **RULE 5:** **Show everyone that the strong men will save them**

The character of 'strongmen' in politics is making a comeback.<sup>98</sup> One of the most prominent examples of this is Vladimir Putin, seeming to restore the 'good old Russia'. In Turkey, Recep Tayyip Erdogan is called 'the tall man' attributing to his physique but also his rule of nearly two decades. In Hungary, Viktor Orban is more powerful than ever. And Donald Trump keeps up with these men; in 2019, he attacked four Democratic congresswomen including Alexandria Ocasio-Cortez, saying they should "go back to where they came from".<sup>99</sup> These 'strongmen' demonstrate similarities in their attitudes towards women. When women try to reclaim their power or raise their voice, they see it as a challenge to their rule as well. Hence they encourage an undermining discourse towards women and other genders, to support their policies such as banning abortion, their anti-LGBT sentiments and their ideas on where women belong.

Comments about Duterte found in the Philippines data divided into positive and negative assessments - both of which employed the idea that a good leader is a strong man. He is seen as "courageous",

"aggressive" and received praise for frequently swearing "putangina" [son of a whore]. Conversely, criticism of him included calling him a "sissy bitch". One tweet said the Philippines would always have a "pussy culture" if it did not honour military personnel (a traditional masculine endeavour). This is in line with how also men who criticise Duterte can be the subjects of gendered harassment - such as Senator Antonio Trillanes, targeted with memes that showed him as "effeminate", and the subject of allegations he was having an affair with a news anchor.<sup>100</sup>

While strongmen are lifted up, women are talked down to. A report published by the International Press Institute is helpful to understand the experiences of female journalists in Poland.<sup>101</sup> They held conversations with female journalists in Warsaw about online attacks against female journalists. Interviewed journalists said they face condescending comments online daily. This includes being addressed by their first name, or a diminutive of it (usually reserved for small children or close family situations), rather than by the polite "Pani" (Ms.) used between strangers in Poland.

As with *Rule Four*, in both the Philippines and Poland datasets we saw 'strongmen' approaches defended on the grounds of protecting women. Bringing back the death penalty, or even the occurrence of extrajudicial killings were justified or called for frequently by the need to prevent rape and punish rapists. This focus on rape, in conjunction with praise for strong men, can create a distorted picture of womanhood, as something primarily characterised by the constant threat of violent attack and therefore in need of protection from those less at risk.

*...I do pity the next President! How, in the world, is he/she's going to get some balls to get things done.*

*feministki boja się silnych i "białych" mężczyzn. reagując agresją jak ktoś się z nią nie zgadza. próbująca ograniczyć kogoś zachowania*

*Feminists are scared of strong and "white" men. Reacting aggressively if somebody doesn't agree with them. Trying to limit their behaviour.*

98. Bremmer, I. The Strongmen era is here: here is what it means for you. Time, 2018. Available at <https://time.com/5264170/the-strongmen-era-is-here-heres-what-it-means-for-you/> [accessed 03 May 20]

99. Pengelly, M. 'Go back home': Trump aims racist attack at Ocasio-Cortez and other congresswomen. The Guardian, 2019. Available at <https://www.theguardian.com/us-news/2019/jul/14/trump-squad-tlaib-omar-pressley-ocasio-cortez> [accessed 5 May 2020]

100. Combinido, P. When Illiberal Social Media Takes over Democratic Philippines. New Mandala, 2019. Available at: <https://www.newmandala.org/when-illiberal-social-media-takes-over-democratic-philippines/> [accessed 5 May 2020]

101. International Press Institute. Online Attacks on Journalists in Poland. International Press Institute, 2018. Available at <https://ipi.media/poland-on-the-line/> [accessed 5 May 2020]

We also saw campaigns aiming to get women in public life currently to leave, by any means possible. In our dataset we saw many calls via hashtags for the Vice President of the Philippines to leave office - either through impeachment, resigning or losing an election. Though some hashtags were used more than others, there was not overall consistency of hashtags used, suggesting perhaps more organic coming up with hashtags or varying on a central theme than attempts to coordinate a campaign via a specific hashtags. Examples included “#OustLeni2020”, “#FakeVP”, “#ImpeachLeniRobredo”, “#LeniResign” and “#GoodbyeLeni”. Some tweets included physically threatening calls to action.

The message to people is clear - swear, be aggressive, and do not, whatever you do, whoever you are, be like a woman. Women are for protecting - they should not be leaders.

**RULE 6:  
Demonise the values women hold**

The amalgamation of Left-wing politics and feminist politics was a central feature of gendered disinformation explored by this research. By eliding the two, conservative or Right-wing governments such as those investigated portray feminism as anti-governmental and oppositional.

In Poland, *Feministka* (feminist, diminutive) was used as a term of abuse. Abuse targeting the left-wing politician Joanna Wicha explained how her shortcomings were the fault of her “leftist, feminist brain”. Advocating for left-wing and centre-left policies were mocked or caricatured through association with feminism: climate change was mockingly attributed to “the patriarchy” in a number of Tweets targeting left-wing Polish politicians. Left-wing positions on abortion and birth control in particular provoked abuse.

*wg lewicy aborcja jest ok bo „zarodek” nic nie czuje. feministka po tabletkce gwałtu też nic nie czuje. więc: -pewnie gwałt się nie liczy.*

*Leftists like you think abortion is fine because an “embryo” doesn’t feel anything. A feminist after a rape can take a morning after pill and also feels nothing, so clearly rape is not a crime either.*

There was little discussion identified of feminism specifically in the Philippines data we examined - however, women were accused of being funded by those opposed to the state (which included anti-semitic tropes): “yellow paid presstitutes” and “yellow funded presstitutes”.<sup>102</sup> There were references to “yellowtards” “yellows” “yellowf\*ckers” “yellowsbits” often in tweets which attacked women - “idiot puppet yellow-ass bitch” or groups in a gendered way such as “presstitutes”. Critics of Duterte were referred to as “yellow trolls”. This reflected a broader trend of women being called “yellows” in a derogatory fashion.

*Why Duterte fired Robredo: One: She is a woman who cannot stop her mouth both up and down Two: She was serious in bringing the government down using her useless dead brain Three: She could not be trusted since Day 1 since she is just an FVP used by the Yellow Crabbies 4. She is becoming too BOBA!*

This dissipates the message that anyone with political views opposing the current regime - and particularly where those views regard the emancipation of women - that those views are illegitimate, disgusting and dangerous.

102. Yellow is associated with the Liberal Party and the political family of former Presidents Corazon Aquino and President Benigno Simeon Aquino, before President Duterte. Before the 2016 elections, the Aquinos and liberals were the targets of demonization on social media.



# CHAPTER 5

## TALKING BACK AND SPEAKING UP

'Women are not passive victims of gendered disinformation and harassment - they are fighting back against it, and have many allies in that fight.'

We used gendered terms to identify where gendered disinformation was occurring in our dataset. A trend which came out of the Philippines in particular, when we examined the data more closely, was that many of these terms are being used by users, including targets of gendered disinformation, to call out their attackers and speak up against abuse. It is important to acknowledge that women are not passive victims of gendered disinformation and harassment - they are fighting back against it, and have many allies in that fight. For instance, Leni Robredo has been a target of sexualized distortion campaigns, including those accusing her of a sexual relationship with a married lawmaker, her hairdresser, and two government officials.<sup>103</sup> False news and social media stories about her "boyfriends" and a supposed pregnancy led her to create a series of Facebook videos addressing each false claim against her, debunking "them one by one".<sup>104</sup>

Here we present an examination of some of the themes of counterspeech we saw most commonly.

Counterspeech is a way to challenge or mock the stereotypes and the norms, to expose the hateful comments, to call out gendered disinformation, or sometimes a way to speak up. In the

Philippines dataset, 1,241 tweets were identified as counterspeech (with a broad understanding of counterspeech as speech which called out or criticised violence against women, whether online or offline). The majority of counterspeech in the Philippines dataset occurred around discussions of "rape" but it was not limited to it. Tweets identified as counterspeech also challenged the gendered disinformation of Duterte critics or speech against LGBT people.

One of the most common themes was challenging the normalisation of culture of rape. There were many tweets criticising the response to rape and how common it is in the society they live in. The tweets both aimed to draw attention to what is not acceptable and how to change it. In August 2018, President Duterte made a comment saying rape is inevitable as long as there were beautiful women.<sup>105</sup> His comment led to large amounts of criticism online. Many original seed users we looked at tweeted against his comment asking him to stop trivialising rape. These tweets also drew attention to the language in reproducing violence. Many of the tweets also refused the idea that introducing harsher penalty to rape or sexual harassment cases will solve the problem without changing the language and culture of rape.

*Sexist words aren't harmless. They enable rape and the culture of sexual harassment. Always, we must stand against this and call out the patriarchy every time that it rears its ugly head. If we remain silent, we are complicit.*

103. Cepeda, M. Here's how Robredo debunks "fake news" vs her. Rappler, 2018. Available at <https://rappler.com/nation/leni-robredo-debunks-fake-news> [accessed 30 July 20]

104. Cepeda, M. Here's how Robredo debunks "fake news" vs her. Rappler, 2018. Available at <https://rappler.com/nation/leni-robredo-debunks-fake-news> [accessed 30 July 20]

105. Filipe Villamor, Duterte Jokes About Rape, Again. Philippine Women Aren't Laughing, The New York Times, 2018. Available at <https://www.nytimes.com/2018/08/31/world/asia/philippines-rodrido-duterte-rape-joke.html> [accessed 15 May 20]

Another common theme in counterspeech was fighting against victim blaming. There were many tweets aiming to educate the audience on why women are not the ones to blame and the importance of consent. Counterspeech happened in the form of speaking against the accepted norms such as a girl should not be drunk or out at night. Many users also discussed what counts as sexual harassment or rape. For instance, a tweet said taking photos of female classmates can not be acceptable or another said the following:

*A girl getting drunk isn't an invitation to molest her. It is not an excuse, for anyone to take advantage of her.*

Countering gendered disinformation was not only limited to discussions around rape. Some of the users speak against gendered insults to undermine people or ideas such as "bakla", "presstitutes" and criticised the use of these terms by mocking or speaking up.

On the other hand, in some tweets, we saw an overlap between content which engaged in counterspeech- speaking against violence against women- while also pursuing the aims of gendered disinformation - invalidating women objecting to rape jokes. For instance, a tweet cited that preventing rape should be a priority rather than being concerned about rape jokes.

Another example that appeared in our dataset, was that of Risa Hontiveros, the opposition Senator who was called a "thirsty slut" on Twitter after a photo of her was shared at a State of the Nation Address wearing a skirt.<sup>106</sup> We found a handful of tweets from original seed users countering this attack, for example, claiming that those who were slutshaming were trying to "belittle or devalue her [...] not just as a woman but someone who has made a massive difference" through her role in government. However, these counter-attacks did not then circulate amongst the friends and followers of original users, suggesting that counterspeech did not necessarily have the same cut-through with wider networks.

106. Rappler. Risa Hontiveros, Twitter users fire back at slut-shaming netizen over SONA dress. Rappler, 2019. Available at <https://www.rappler.com/nation/236011-risa-hontiveros-twitter-users-fire-back-slut-shaming-netizen-sona-2019-dress> [accessed 15 May 20]

# CHAPTER 6

## THE GENDERED DISINFORMATION ECOSYSTEM

**'Gendered disinformation online does not occur in a vacuum.'**

Gendered disinformation online does not occur in a vacuum. It is linked to what is happening in the news, what commentators are saying, what politicians are saying, and varies across platforms as user bases and platform features vary the kind of content that is most prominent.

Our analysis centered on Twitter, as the most accessible platform for programmatically collecting large amounts of data from; however discourse on Twitter does not always represent political discourse as a whole in a country. For instance, in the Philippines, Facebook is a particularly relevant site of political discourse. Hence to get a better picture of the broader landscape, we examined URLs that were shared within tweets identified as being relevant to gendered disinformation.

In Poland, a considerable number of the users and followers were sharing content from media outlets considered mainstream or traditional, such as DoRzeczy, TVN24, wPolityce, Interia and Onet.<sup>107</sup> A smaller number, though, were sharing news articles from more overtly partisan sites such as wPrawo, an online media site set up by the former priest and far-right activist that pushes nationalist and anti-Ukrainian themes and Natemat, described

as "a forum for left-wing and liberal commentators" against the ruling party PiS.<sup>108</sup>

Outside of those sharing news outlets, tweets would share content from political campaigns, across the political spectrum. A further handful of tweets from friends and followers would directly troll women politicians and a smaller number still directly endorse them.

In the Philippines, content shared by original users consisted mainly of two approaches to disinformation: either sharing or endorsing mainstream news sites and social media news sites; or directly retweeting content that might be considered empowering with sexual distortion and hate speech to discredit and devalue it. For example, when politician Leila De Lima received an award from the Financial Times for her worldwide influence, a pro-Duterte account shrugged this off by sexualising her role in politics as simply "just like masturbation, stock prices go up when Leila De Lima says saba".

A Facebook page, run by someone internationally recognised as a pro-Duterte proponent of gendered disinformation, was also regularly shared by users with content on key female figures such as Leila De Lima.<sup>109</sup> The account owner has a social media presence as well as a media presence.

107. Bush, D., Gielewska, A., and Kurzynski, M. Polish Presidential Election 2020: Two Months Out, Stanford Internet Observatory, 2020. Available at <https://cyber.fsi.stanford.edu/io/news/polish-elections-2-months-out> [accessed 14 April 2020]

108. Eurotopics.net. Natemat. Available at <https://www.eurotopics.net/en/148727/natemat> [accessed 14 April 2020]

109. Etter, L. What Happens When the Government Uses Facebook as a Weapon?, Bloomberg Businessweek, 2017. Available at <https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook> [accessed 13 April 20]

A minority of the content shared by original users was to counter gendered disinformation, including the targets themselves sharing content e.g. about Duterte “slut-shaming” them.

In the Philippines, the shared content from the friends and followers of the original users echoed that of the original users. The content shared by friends and followers in particular included a variety of news sites and Facebook pages that were overtly partisan.

These links being shared between Facebook and mainstream news sites on Twitter suggest that what is happening on Twitter is indeed linked into and can be used to obtain a snapshot, even though not a full representation, of the broader disinformation ecosystem.<sup>110</sup>

110. Zeiter, K., Pepera, S., and Middlehurst, M. Tweets That Chill: Analyzing Online Violence Against Women in Politics. National Democratic Institute, 2019. Available at <https://www.ndi.org/sites/default/files/NDI%20Tweets%20That%20Chill%20Report.pdf> [accessed 03 May 20]

# CHAPTER 7

# CONCLUSIONS

'These rules, used together, build up a toxic environment online where, if women dare to enter, they are abused, undermined, threatened, gaslit, delegitimised, and insulted.'

Taken together, what are the aims and methods of state-aligned gendered disinformation, and what does this mean for the targets, for politics and for policy?

These rules, used together, build up a toxic environment online where, if women dare to enter, they are abused, undermined, threatened, gaslit, delegitimised, and insulted. Not only can this cause serious harm for individuals involved, but creates a public sphere in which women are not free to express themselves, or to participate in politics, or to speak up against injustice and oppression. These stories have important commonalities in how they seek to attack and undermine women in public life. We conclude by outlining strategies which bad actors are employing to promote political outcomes through the manipulation of gendered norms.

## **THEY USE EMOTIONS, NOT FACTS**

The clearest commonality between the stories told about women in public life described above is that they seek to inflame people's emotions, in addition to spreading false information about their targets. While this includes false information, the content of the messages is clearly not only seeking to make assertions but to anger those reading it - by playing on highly emotive, value-based concepts such as loyalty, honesty, purity; or to persuade them to dismiss someone as laughable by painting them as stupid, weak and emotional. Fear is also

weaponised - fear of violence, fear of mysterious conspiracies, or even fear of their own safety - in order to emotionally manipulate readers.

## **THEY TARGET WOMEN, CONVERT THEIR ALLIES, BOLSTER THEIR CRITICS**

Online discussions criticising women in public life in the ways set out above also have two targets. One is the woman herself - tweets which either directly tag their Twitter handle and address them, or which speak in the second person when discussing them. Other discussions are speaking either to other parties about their targets; expressing an opinion generally; supporting their critics or attacking their supporters, by presenting gendered reasons why women should not be believed or trusted to lead. The messages work to engender anger, disgust and disdain in the third party reader; and fear and shame in the second person target.

## **THEY USE IN-JOKES**

Specific phrases or terms are used in online discourse by communities engaging in SAGD to denigrate a particular woman or make reference to a particular event, which might not be immediately obvious to an outside observer. For instance, many tweets attacking Senator Leila de Lima did not mention her by name (and so would not be found if looking through tweets about her by name alone) but used the term "saba queen" instead (referencing the rumours about her having an affair). These in-group usages mean that it is easier to identify who is 'on your side' and to find other like-minded people in a formal or informal network who are likely to support your aims. Hashtags are also used to similar effect.

## **THEY MAKE THE NEWS - THEN USE THE NEWS**

We saw online discussion around notable instances of gendered disinformation - abuse or offensive

comments made by political leaders, which then became a discussion point in themselves which led to repetition online, including in mainstream media

### **THEY ARE PARASITIC**

We saw gendered disinformation and terms relevant being used to reinforce social norms and gender stereotypes against critics and supporters of the state. This shows that disinformation relies on co-opting underlying values, beliefs and ways of speaking to get its message across, rather than trying to invent something completely new, which can then result in methods of disinformation spreading more widely and being used against people beyond the original targets.

# CHAPTER 8

## RECLAIMING THE GAME

**'Identifying or mitigating gendered disinformation cannot be successful without the central involvement and direction of local experts.'**

In this report we present a methodology which can be used to identify where state-aligned gendered disinformation is occurring in an online environment, and what forms it takes and purposes it aims to serve.

To challenge the harms that are caused by state-aligned gendered disinformation, we need platforms, policymakers and civil society to recognise that gendered disinformation exists as a specific form of gendered online violence and of disinformation, which requires its own tailored interventions.

This is not to say that nothing is being done. As we have demonstrated in this report, women are speaking up against these campaigns, and politicians, journalists and activists are continuing to do their vital work in spite of these efforts to stop them. However, the onus should not be on women who are being targeted in these campaigns to deal with it themselves. Being able to put up with coordinated campaigns of hatred, abuse, lies and misogyny should not be a requirement for being a woman in public life.

It is of course crucial that users are given the controls on the platforms they engage in to block or report users or content, or to limit visibility of or engagement with their own content, or to

delete their accounts - all of which women use to respond to online violence. Moreover, platforms must actually follow through in responding to such reports.<sup>111</sup> There is more work that can be done to enforce this, from ensuring the enforcement of terms of service to making reporting mechanisms more transparent, easy to use and accountable.<sup>112</sup> But assuming that women will deal with the problem themselves, on a comment-by-comment basis overlooks the systematic effects of gendered disinformation, as well as putting an unfair burden on women who are not only being targeted for what they say, but as we have shown, also for responding to or blocking people on the basis of gendered disinformation.

We need systematic improvements to be made to how platforms operate in facilitating, promoting and moderating online speech, and how policy and practice respond to these problems. We present here the challenges of existing solutions to these problems, and criteria that effective solutions must meet. Thanks must go to the participants of a closed roundtable held by NDI to discuss these issues in helping inform these recommendations.

### **RECOGNISE THE DIVERSITY OF DISINFORMATION**

Disinformation takes many forms and is engaged in and perpetuated by many different actors: a 'one-size fits-all' approach will not be successful. Locating the harm of gendered disinformation as inherent in the fact that what is being said is untrue is to misunderstand the violent and prejudicial elements of what is being said. Hence fact-checking, while important in the fight against

111. Iyer, N., Nyamwire, B., and Nabulega, S. *Alternate Realities, Alternate Internets*. Pollicy, 2020. Available at <https://ogbv.pollicy.org/report.pdf> [accessed 24 August 2020]

112. Thanks to participants at a closed roundtable with NDI for making this point. August 2020.

disinformation, is not a panacea, particularly where the claims being made are couched in value judgements or are about people's character.

Similarly, there is a spectrum of disinformation: from the huge, clearly identifiable campaigns involving clearly inauthentic accounts, down to low-scale attacks from a small group of like-minded individuals. Focusing on dealing with the problem at both of these ends as the easiest to identify and act on will miss engaging with the problem of the informal networks we have shown are implicated in gendered disinformation: the political influencers and their followings.

### **NOT JUST A 'GENDER' PROBLEM**

Steps should be taken to tackling gendered disinformation within broader programmes of work to reduce gender inequality, reduce gender-based violence, and promote equal democratic participation. However, tackling gendered disinformation should not be the sole purview of gender-based programming: broader responses to online harms, analysis of information operations, or improvements to platform design and oversight, should take into account the ramifications for the spread or encouragement of gendered disinformation.

### **CENTRE LOCAL EXPERTISE**

This report has shown that though the narratives perpetuated may be similar, how gendered disinformation is framed and spreads across a network varies greatly according to context. Identifying or mitigating gendered disinformation cannot be successful without the central involvement and direction of local experts

**'Solutions to gendered disinformation will not succeed if they are imposed upon the women who are being most affected.'**

who understand the subtleties of how gendered disinformation may be expressed and where it is likely to arise and when. Platforms should support the work of local experts in identifying and combating gendered disinformation, for instance through the provision of data access or the trialling of potential responses through changes to platform design. Automated systems for identifying gendered disinformation are unlikely to have high levels of accuracy - though if employed, should be employed transparently and overseen by local experts.

### **ELEVATE WOMEN'S VOICES**

Solutions to gendered disinformation will not succeed if they are imposed upon the women who are being most affected: indeed, it is these women who have experienced the positives and negatives of existing solutions most significantly. Solutions should be co-created with people who have experienced being the targets of gendered disinformation, to ensure that the most significant threats are being addressed in a way that does not compound the problem. Metrics of success used should be rooted in how positive and empowering spaces for women to participate in have been built, rather than simply how much bad content has been taken down.<sup>113</sup>

113. See e.g. Iyer, N., Nyamwire, B., and Nabulega, S. *Alternate Realities, Alternate Internets*. Policy, 2020. Available at <https://ogbv.policy.org/report.pdf> [accessed 24 August 2020]



# DEMOS

PUBLISHED BY DEMOS OCTOBER 2020

© DEMOS. SOME RIGHTS RESERVED.

15 WHITEHALL, LONDON, SW1A 2DD

T: 020 3878 3955

HELLO@DEMOS.CO.UK

WWW.DEMOS.CO.UK