unesco

The Chilling: Global trends in online violence against women journalists

Research Discussion Paper



Authors:
Julie Posetti
Nabeelah Shabbir
Diana Maynard
Kalina Bontcheva
Nermine Aboulez

Content warning: This document includes graphic content that illustrates the severity of online violence against women journalists, including references to sexual violence and gendered profanities. This content is not included gratuitously. It is essential to enable the analysis of the types, methods and patterns of online violence.

Disclaimer: The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The views and opinions expressed in this document are those of the authors and should not be attributed to UNESCO.

This discussion paper is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (http://creativecommons.org/licenses/by-sa/3.0/igo/). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (http://en.unesco.org/open-access/terms-use-ccbysa-en).

Graphic Design: Dean Dorat

Cover and illustrations: Franziska Barczyk

Published by UNESCO, April 2021.

CI-2021/FEJ/PI/1

About this research discussion paper:

This paper is an output of a wider UNESCO-commissioned global study on online violence against women journalists, with a full-length study to be published in mid-2021. As a means of further informing the forthcoming study, the paper seeks to promote discussion and elicit information about effective legislative, organisational and normative initiatives, aligned to international standards for freedom of expression, which are designed to protect women journalists.

ICFJ online violence project research team:

Lead researcher: Dr. Julie Posetti. Global Director of Research, International Center for Journalists (ICFJ); Senior Researcher, CFOM, University of Sheffield; Research Associate RISJ, University of Oxford.

Senior Research Associate: Nabeelah Shabbir. ICFJ researcher and freelance journalist.

Research Associate: Nermine Aboulez, ICFJ researcher and University of Oregon PhD candidate.

Specialist researchers: Prof. Kalina Bontcheva, Angelique Lu, Dr. Diana Maynard. **Africa:** Assoc. Prof. Glenda Daniels (Regional Lead); Fiona Chawana, Dr. Omega Douglas; **Arab States**: Nermine Aboulez (Regional Lead); **Asia and the Pacific:** Assoc. Prof. Fiona Martin (Regional Lead), Liana Barcia, Ayesha Jehangir, Nirasha Piyawadani, Dr. Jenna Price; **Central and Eastern Europe:** Dr. Greta Gober (Regional Lead), Jen Adams, Bojana Kostić; **Europe and North America:** Dr. Julie Posetti (Regional Lead), Prof. Jackie Harrison, Nabeelah Shabbir, Dr. Sara Torsner, Prof. Silvio Waisbord; **Latin America:** Dr. Luisa Ortiz Pérez (Regional Lead), Kate Kingsford, Carolina Oms, Kennia Velázquez.

UNESCO editorial coordination:

Saorla McCabe, Theresa Chorbacher.

Project support: Jen Adams, Fatima Bahja, Heloise Hakimi Le Grand, Senka Korać, Sharon Moshavi, Eunice Remondini, Erin Stock, Mengyang Zheng (ICFJ); Lou Palin, Johann Bihr, Sara Bonyadi (UNESCO).

Project Partners: Centre for Freedom of the Media (CFOM), University of Sheffield; Dart Asia Pacific; Ethical Journalism Network (EJN); International Association of Women in Radio and Television (IAWRT).

This project has received financial support from UNESCO's Multi-Donor Programme on Freedom of Expression and Safety of Journalists and the Swedish Postcode Foundation.

The easiest part is dealing with the impact of online violence and disinformation on me. I just see the impact on the world, and I don't know why we're not panicking.

Maria Ressa[1]
Rappler CEO and co-founder
Laureate of the 2021 UNESCO/Guillermo Cano World Press Freedom Prize

[1] Maria Ressa is an internationally celebrated Filipino-American journalist whose situation is the subject of a major interdisciplinary case study highlighted in this report.

Introduction

There is nothing virtual about online violence. It has become the new frontline in journalism safety - and women journalists sit at the epicentre of risk. Networked misogyny and gaslighting intersect with racism, religious bigotry, homophobia and other forms of discrimination to threaten women journalists - severely and disproportionately. Threats of sexual violence and murder are frequent and sometimes extended to their families. This phenomenon is also bound up with the rise of viral disinformation, digital conspiracy networks and political polarisation. The psychological, physical, professional, and digital safety and security impacts associated with this escalating freedom of expression and gender equality crisis are overlapping, converging and frequently inseparable. They are also increasingly spilling offline, sometimes with devastating consequences.

Here, we present an edited extract from a major interdisciplinary study produced by the International Center for Journalists (ICFJ) under commission from UNESCO. The book-length study will be published by UNESCO in mid-2021.

The research underpinning this paper consists of: a global survey of **901 journalists**[2] from **125 countries** conducted in five languages[3]; long-form **interviews**[4] **with 173 international journalists, editors, and experts** in the fields of freedom of expression, human rights law, and digital safety; **two big data case studies** assessing over **2.5 million posts on Facebook and Twitter** directed at two prominent women journalists (Maria Ressa in the Philippines and Carole Cadwalladr in the UK) undertaken to validate the self-reporting of our interviewees and survey respondents with objective data; **15 detailed country case studies**[5]; and a literature review covering hundreds of scholarly and civil society research publications. A team of **24 international researchers**[6] **from 16 countries** contributed to the study.

One of the unique aspects of this research is its **focus on understudied developing countries** recognising that online violence against women journalists is a global problem, but one with disproportionate offline impacts and complex intersectional challenges that inhibit effective responses. Our **interviewees represent 17 States** and span every UNESCO region. They are **racially, ethnically and linguistically diverse**. They also include a number of men (e.g., senior editorial leaders, digital security, legal and freedom of expression experts), **people of different abilities**, and they **express a range of sexualities and gender identities**.

[2] 714 respondents identified as women. All survey results presented here are expressed as a proportion of this sample unless otherwise indicated.
[3] Arabic, English, French, Portuguese and Spanish.
[4] All quotes featured in this report are from original interviews unless otherwise indicated.
[5] Kenya, Nigeria, South Africa, Lebanon, Tunisia, Pakistan, the Philippines, Sri Lanka, Poland, Serbia, UK, US, Sweden, Brazil and Mexico.
[6] In addition to the identified authors of this report, the researchers named on the inside cover can choose to be recognised as contributing authors.

The chilling effect

Online violence against women journalists is designed to: belittle, humiliate, and shame; induce fear, silence, and retreat; discredit them professionally, undermining accountability journalism and trust in facts; and chill their active participation (along with that of their sources, colleagues and audiences) in public debate. This amounts to an attack on democratic deliberation and media freedom, encompassing the public's right to access information, and it cannot afford to be normalised or tolerated as an inevitable aspect of online discourse, nor contemporary audience-engaged journalism. As UN Secretary-General António Guterres tweeted in March 2021 in connection with this research:

> ***There should be no room for misogyny and violence in journalism. Social media platforms and governments must protect women journalists from online violence.***
>
> António Guterres
> UN Secretary-General
> March 2021

A worsening crisis

Online attacks on women journalists appear to be increasing significantly, as this study demonstrates, particularly in the context of the 'shadow pandemic' of violence against women during COVID-19. The pandemic has changed journalists' working conditions, making them yet more dependent on digital communications services and social media channels. The emergence of the 'disinfodemic' has also increased the toxicity of the online communities within which journalists work, making journalists "sitting ducks" according to the UK National Union of Journalists' Michelle Stanistreet, interviewed for this study.

Women journalists are now more exposed to online violence than ever.

Our research also highlights the threefold function of disinformation in gendered online violence against women journalists:

1 // Disinformation tactics are routinely deployed in targeted multiplatform online attacks against women journalists.

2 // Reporting on disinformation and intertwined issues, such as digital conspiracy networks and far-right extremism, is a trigger for heightened attacks.

3 // Disinformation purveyors operationalise misogynistic abuse, harassment and threats against women journalists to undercut public trust in critical journalism and facts in general.

Another major issue in evidence is the role of political actors - including presidents and elected representatives, party officials and members - in instigating and fuelling online violence campaigns against women journalists. Additionally, partisan, mainstream and fringe news media can be shown to amplify such attacks, triggering 'pile-ons' that escalate the risks of online violence morphing into offline assault or causing significant psychological injury.

While the platforms are the main vectors or enablers for online violence against women journalists, media employers are responsible for ensuring a safe working environment for their journalists. And States are obliged to protect journalists, so that they can do their work freely and safely in accordance with international obligations, legislating accordingly and ensuring law enforcement agencies respond appropriately.

Problematically, however, women journalists are both the primary targets of online violence and the first responders to it. On the unmediated social media platforms they use in the course of their work, they are the ones required to 'report', 'block', 'mute', 'delete', and 'restrict' their attackers, potentially compounding the effects of the abuse, and creating unbearable pressures when the attacks come at scale.

Online violence against journalists is a significant feature of what we call 'platform capture', which involves the weaponisation of social media by bad-faith actors, in combination with the structural failures of the platforms' business models and product design, and the virtual entrapment of many news organisations and journalists into platform dependency.

While making fledgling attempts to improve their products to enable easier reporting and filtering of abuse, and engaging in regular public relations exercises designed to promote their commitment to safer platforms, our research indicates that the social media companies have so far failed to respond swiftly or effectively to the crisis. Central to this failure is an attempt to use 'free speech' as a shield against accountability, and a continuing reluctance to assume responsibility for the content on their sites. Additionally, there is a lack of coordination between these companies in addressing orchestrated, cross-platform online violence campaigns targeting individual journalists. There have also been few attempts to hold social media companies liable or make them accountable through legal redress or forms of regulation.

Media employers have made some progress in addressing gendered online violence against their journalists (primarily within the comments sections of their websites), but the journalism safety threats posed in the networked social media environment, and the risks intersecting with disinformation and political extremism (particularly on the far-right), appear to be either poorly understood or too overwhelming to manage in many cases. Increasingly, and very problematically, employers respond to the problem by policing journalists' speech (e.g., by introducing social media policies that discourage them from engaging in public commentary on "controversial issues") and victim-blaming (e.g., by suggesting a woman's speech triggered an attack, or punishing them for the brand exposure caused by an attack).

Women journalists are both the primary targets of online violence and the first responders to it.

Simultaneously, social media companies, news organisations and States are struggling to respond effectively to online violence against women journalists. This highlights the urgent need for policy reform, and novel legal, legislative and normative responses, in full compliance with international freedom of expression and privacy standards, from key stakeholders - in particular UNESCO Member States. This research discussion paper - published ahead of the major forthcoming study - seeks to help stimulate, identify and elicit such responses.



Defining and describing online violence

Gendered online violence can be understood as a combination of:

// **Misogynistic harassment, abuse and threats.**

// **Digital privacy and security breaches that increase physical risks associated with online violence.**

// **Coordinated disinformation campaigns leveraging misogyny and other forms of hate speech.**

Online violence targeting women journalists manifests itself in a variety of ways, but it has a number of common characteristics:

// It is networked

Online violence is often organised, coordinated or orchestrated. It can include State-sponsored 'sock puppet networks,' acts of 'patriotic trolling,' and involve mobs who seed hate campaigns within one fringe network before pushing it into more mainstream networks and partisan media. But such abuse can also come from individuals united in a common cause - like misogyny.

// It is usually misogynistic

Misogyny is one of the key features of online violence targeting women journalists, and it has been routinised.

// It radiates

The perpetrators of online violence against women journalists often target their families, sources, colleagues and bystanders too.

// It is intimate

In detail and delivery, the threats are personal. They arrive on mobile phone screens first thing in the morning and last thing at night, and they are often highly sexualised.

// It can be extreme, intense and prolific

Often described by targets in terms of extreme weather events, natural disasters and war, such as: "torrential", "tsunami", "flood", "avalanche", "a barrage", "trench warfare", "bombardment".

// It can behave like networked gaslighting

Constant moderate-low volume abuse and harassment that burns slowly but can be cumulatively devastating.

Top 26 Preliminary Findings

Manifestations of online violence

1 **Nearly three quarters (73%)[7] of our survey respondents identifying as women said they had experienced online violence.**

2 **Threats of physical violence (identified by 25% of survey respondents) including death threats, and sexual violence (identified by 18%) also plagued the women journalists we interviewed. And these threats radiated: 13% of survey respondents and many interviewees said they had received threats of violence against those close to them, including children and infants.**

3 **One-fifth (20%)[8] of survey respondents identifying as women said they had been attacked or abused offline in connection with online violence they had experienced. A similar proportion of our interviewees also experienced offline abuse associated with online attacks, including the subjects of both of our big data case studies.**

4 **Racism, religious bigotry, sectarianism, ableism[9], homophobia and transphobia intersect with misogyny and sexism to produce significantly heightened exposure and deeper impacts for women experiencing multiple forms of discrimination concurrently, as evidenced by our survey respondents and interviewees, and detailed in our big data case study on Maria Ressa.**

5 **Black, Indigenous, Jewish, Arab and lesbian women journalists participating in our survey and interviews experienced both the highest rates and most severe impacts of online violence.**

6 **Online violence against women journalists verges from large-scale attacks or extreme threats at a moment in time, through to the slow burn of networked gaslighting, which involves constant lower level abuse.**

[7] Expressed as a percentage of 625 woman-identifying respondents who answered this question.
[8] Expressed as a percentage of 596 woman-identifying journalists who responded to the question.
[9] Discrimination in favour of able-bodied people i.e. those who do not have disabilities.

Impacts of online violence

7 Physical threats associated with online violence caused 13% of women survey respondents to increase their physical security; 4% said that they had missed work due to particular concerns about the attacks moving offline and resulting in physical violence. This pattern was also evident among the interviewees, several of whom relocated due to the physical threats associated with online attacks.

8 A number of our interviewees were suffering from PTSD connected to online violence, and many were in therapy as a result. The mental health impacts were also the most frequently identified (26%) consequence of online attacks among survey respondents. 12% of respondents said they had sought medical or psychological help due to the effects of online violence.

9 When asked "How does the level of online violence you experience affect your journalism practice and your interaction with sources/audiences?", 30% of the women journalists surveyed answered that they self-censored on social media. 20% described how they withdrew from all online interaction. Self-censorship was also a response noted by many interviewees.

10 Employment and productivity impacts reported by the women survey respondents included missing work to recover from online violence (11%), making themselves less visible (38%), quitting their jobs (4%), and even abandoning journalism altogether (2%). Linked to this was the professional discreditation of online violence targets. The interviewees confirmed this pattern, and the big data case study on Carole Cadwalladr potently underlines it.

Online violence sources and triggers

11 The reporting theme most often identified in association with heightened attacks was gender (49%), followed by politics and elections (44%), and human rights and social policy (31%), according to our survey. This was also borne out by our interviews and case studies.

12 41% of survey respondents said they had been targeted in online attacks that appeared to be linked to orchestrated disinformation campaigns. This trend was underlined by the interviewees and also linked to reporting on far-right extremism and conspiracy networks. Additionally, it was highlighted by both of our big data case studies.

13 Our interviewees and case study subjects identified high-level political leaders and other State actors as some of the biggest instigators of online violence. Political actors were also the second most frequently noted sources (37%) of attacks and abuse after 'anonymous or unknown attackers' (57%), according to the women survey respondents.

14 Partisan news outlets, media operating at the fringes of the political spectrum, and misogynistic journalists are instrumentalised to amplify and fuel attacks. This is a trend identified across our interviews and case studies.

15 Online violence against women journalists has worsened in the context of the COVID-19 pandemic.

The platforms as vectors of online violence

16 Social media companies are the main enablers of online violence against women journalists.

17 Facebook was the most frequently used (77%) platform, network or app for journalistic work (closely followed by Twitter on 74%) according to the women surveyed, but it attracted disproportionately higher rates of incident-reporting to the platforms among the respondents (39% compared to Twitter's 26%). It was also rated the most dangerous of the top five platforms/apps used, with nearly double the number of respondents rating Facebook "very unsafe" compared to Twitter.

18 Despite fledgling responsive efforts and stated commitments to enhancing journalists' safety on their platforms, the social media companies are failing to stem online violence against women journalists, with sometimes devastating consequences, as illustrated through our big data case studies and interviews.

19 In addition to technical design and business model failures that enable online violence on their platforms, social media companies lack gender-sensitive, human-focused solutions to online attacks on women journalists. Despite capacity to do more, they also lack rapid response units and capability to respond in all languages that their platforms allow.

20 Online violence frequently jumps platforms in networked attacks, demanding networked responses.

The role of media employers as responders to online violence

21 **Despite progress made by many employers over the past five years, only 25% of the women survey respondents said they had reported online violence incidents to their employers, and the top responses they said they received when they did were: no response (10%) and advice like "grow a thicker skin" or "toughen up" (9%), while 2% said they were asked what they did to provoke the attack.**

22 **Evidence of victim-blaming by employers is accompanied by the emergence of a pattern of policing women journalists' speech as a response to 'managing' online violence through more restrictive and punitive social media policies introduced by news organisations.**

23 **In some cases, women journalists are actively discouraged from speaking about their experiences of online violence or engaging with those attacking them. They are also sometimes told to avoid discussing "controversial" topics on social media as a preventive measure. In the worst cases, they have been suspended or sacked from duty in the midst of a major attack.**

24 **A range of obstacles inhibits more effective responses to online violence by news organisations. These include: external factors such as social media companies' failure to adequately address platform design and business model flaws that enable online violence and empower abusers.**

Legal remedy and law enforcement

25 **Only 11% of the women journalists surveyed had reported instances of online violence to the police, and very few interviewees had made a complaint to law enforcement agencies, highlighting a general lack of confidence in prosecution.**

26 **Only 8% of the women journalists surveyed and just a few of our interviewees had taken legal action, highlighting a reluctance among those targeted to pursue legal remedy due to various impediments. This signifies the need for improvement in legal and judicial responses to online violence against women journalists. Two of our interviewees[10] were pursuing legal action against State actors (foreign and domestic) at the time of writing.**

[10] They are: Lebanese journalist Ghada Oueiss and Brazil's Patricia Campos Mello.

Global thematic analysis: key trends



We synthesised the data from the 173 original research interviews, the 15 country case studies and the gender-disaggregated results from our survey of 901 journalists, along with the two big data case studies detailing the online violence experienced by Maria Ressa and Carole Cadwalladr, to conduct an analysis that led us to identify eight key international trends in online violence against women journalists. They are:

173
research interviews

901
survey respondents

1 / Intersectional threats

Women journalists who are also disadvantaged by forms of discrimination that intersect with sexism and misogyny (e.g., racism, homophobia, religious bigotry, sectarianism) face additional exposure to online attacks, with worse impacts.

2 / Online violence moving offline

There is increasing evidence that online violence against women journalists moves offline with significant impact. This includes physical attacks and offline abuse and harassment that is seeded online, as well as legal harassment enabled and reinforced by online violence.

3 / Disinformation as a multi-pronged and intersecting threat

The weaponisation of false and misleading content functions both as a method of attack (e.g., the deployment of disinformation tactics) and a lightning rod for attacks (stimulating misogynistic 'pile-ons') in the context of online violence against women journalists. Simultaneously, orchestrated disinformation campaigns operationalise gendered online violence to chill critical reporting.



2.5m
social media posts analysed

15
country case studies

survey respondents from 125 countries

4 / The role of political extremism (in particular far-right extremism), nationalism and populism

Misogyny is weaponised in the global tilt towards populism, and women journalists are clear targets - particularly those reporting on far-right extremist networks.

5 / The platforms as vectors of online violence

Social media platforms are seen as the major enablers for online violence against women journalists but cast as (largely) failed responders to the problem.

6 / Political actors as attackers

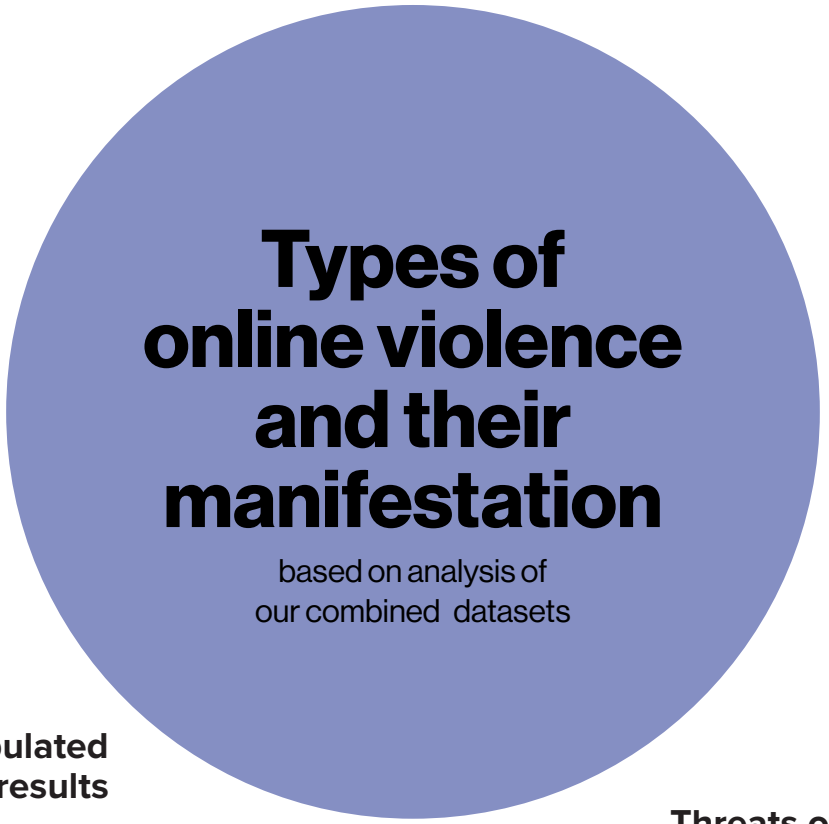
Politicians and political party officials/donors are implicated as major instigators and amplifiers of online violence against women journalists.

7 / Instrumentalisation of partisan media

Partisan media outlets, and mainstream media actors sympathetic to certain political ideologies, are identifiable as instigators and amplifiers of online violence against women journalists.

8 / News organisations still struggling to respond effectively

There is a need to reboot threat modelling and gendered online violence response protocols. This is in the context of an increasingly toxic information ecosystem - in particular at the intersection of disinformation/ far-right extremism and diverse forms of discrimination. In parallel, there is an alarming trend involving employers victim-blaming and policing women journalists' speech when they are targeted.



Misrepresentation

via spoof accounts and manipulated images, video and audio. Such synthetic media were shared openly via social media and on fringe sites that facilitate coordinated attacks, or via email, text message etc.

Threats to damage their professional or personal reputations

Coordinated 'dogpiling'[11] attacks

involving not just fringe networks of misogynists and disinformation agents, but also publicly identifiable political actors and male journalists

Digital privacy and security-based attacks

including hacking, doxxing, interception and surveillance

Orchestrated attacks involving State actors

Harassment in the form of fraudulent online orders

being sent to women journalists' homes after doxxing

Manipulated search results

on YouTube and Google which were flooded with hateful and/or disinformation content designed to malign and discredit the targets, drowning out their professional journalistic content

Harassing private messages

(including text messages) which often involved soliciting for sex

Threats of sexual assault and physical violence

including death threats (some of which were extremely graphic). These were generally received via open social media platforms, closed messaging, email, and text messages. Some journalists also received death threats that came via analogue means (e.g., traditional post) after being doxxed

[11] Dogpiling' refers to a way of loosely organising targeted attacks against women journalists online through digital 'dog whistling'. It is related to 'brigading', which is synonymous with the term 'piling-on'. All these terms refer to the function of online mobs called to attack an individual.

i. WEAPONISING MISOGYNY

Threats of physical and sexual violence plagued the women journalists we interviewed and surveyed. Our interviewees testified to experiencing some particularly chilling threats. One example among many came from Lebanese journalist **Ghada Oueiss**[12] , Al Jazeera's principal Arabic presenter, who has been the target of a prolific online violence campaign focused on her gender, her Christianity, and her age. She was portrayed as a prostitute by her attackers, and she told us:

> ***"[Every day I went on air], I would receive on my Al Jazeera email - because somehow it was leaked - a death threat. One of them that I can never forget [said]: 'You will be looking at the camera to talk to your audience and you will start reading the bulletin and reading the autocue in front of you. You will notice that there is a gun and [a] bullet, that bullet will go straight to your head.' Then I started getting emailed pornographic pictures... they started to put my head on naked women. And then, they made another email in my name and they started sending to my colleagues pictures of [my] head on a naked body, also porn pictures."***

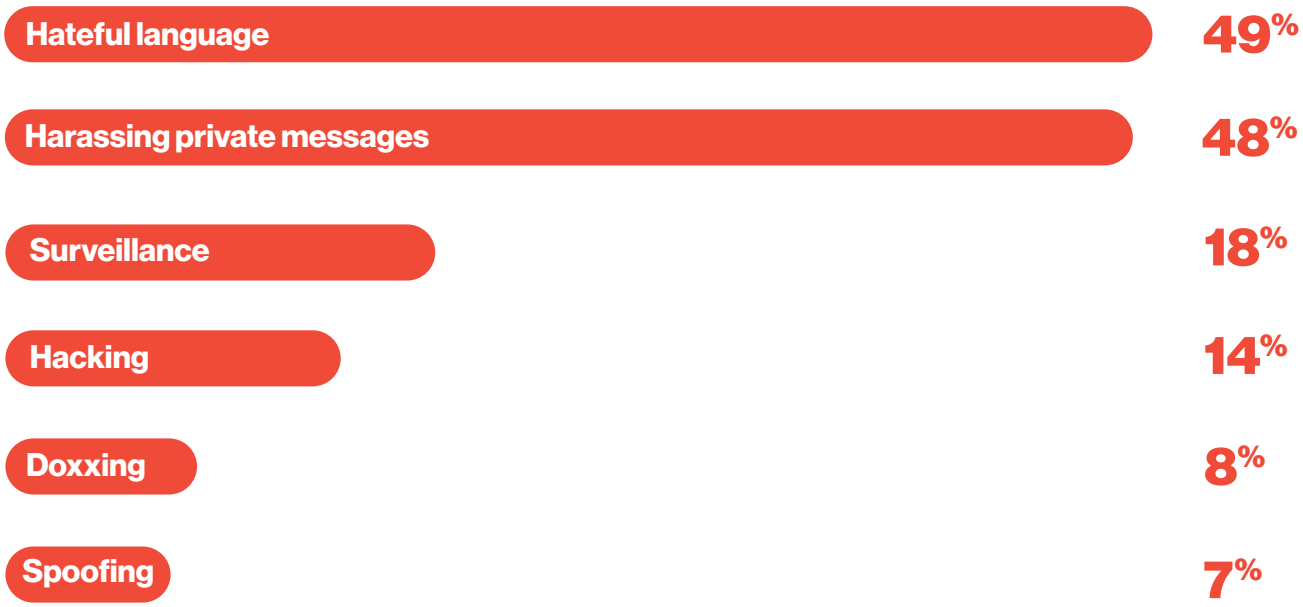
[12] Oueiss is currently pursuing legal action against various officials and individuals in relation to the coordinated online attacks she experiences.

Bystanders who rallied to the defence of our research participants - sources, colleagues, and audiences - along with family members (including children), were also targeted. In fact, 13% of our survey respondents said that in the course of online violence episodes, they had received threats of violence against those close to them.

Caoilfhionn Gallagher QC represents dozens of women journalists from the BBC Persian service who have experienced prolific online violence suspected to be connected to a foreign State actor. In two cases, children of the women were also targeted. In the first instance, the head of one of the presenters whom Gallagher represents was Photoshopped onto a pornographic image, which was then sent to her 14-year-old son at his school:

> *“[There] are multiple examples from social media platforms - Twitter, Instagram - which are horrifying. They are rape threats. They’re horrendous threats of sexual violence either to the journalists themselves, or to family members. And sometimes they’re misogynistic threats to male journalists singling out female family members, or false allegations about male journalists raping their female colleagues.”*

Online threat experience



Almost half (48%) of the women journalists we surveyed reported being harassed - often sexually - with unwanted private messages, highlighting the fact that much online violence occurs in the shadows of the internet, away from public view. But the most commonly reported (49%) online threat experienced by the women surveyed and interviewed came in the form of abuse laced with hateful language designed to denigrate them. This abuse was most noticeably misogynistic and sexist. Classic tropes identified by our interviewees, and evident within our big data collections, included: "witch", "hag", "whore", "bitch", and "presstitute" (a portmanteau of 'press' and 'prostitute').

Dozens of our interviewees had been doxxed - several had their addresses emails and phone numbers shared on extremist websites, and sites designed to facilitate the networked trolling of women. A number - in the US, Sri Lanka, the UK and South Africa - had to move house, region or even country. Digital security and privacy attacks against the women survey respondents included reported exposure to surveillance (18%), hacking (14%), doxxing (8%); and spoofing (7%). This is a significant set of findings that not only indicates increased susceptibility to physical violence for the women involved (because of the public identification of personal details, including physical addresses), but also increased threats for at-risk confidential sources, and their family members.

Al Jazeera's White House correspondent, **Kimberly Halkett** - a work-related sexual assault survivor - was targetted in a brutal campaign of online violence in 2020 after being doxxed:

13%

of women survey respondents received threats of violence against those close to them, in the course of online violence episodes

> ***"They wanted to kill me. They were going to come after my family. My address was posted all over the internet. But the part that really upset me and that rattled me to my core is when they went after my 15-year-old daughter... .They found a post where I talked about her recently doing well in math and I had tagged her, and so they began to blow up her Instagram, saying all kinds of horrible things about her, about me... My daughter had done nothing except be my daughter."***

Finally, nine percent (9%) of our survey respondents reported being threatened financially, including attempts to extort money or assets, hacking into bank accounts, and misrepresentation to financial institutions. In Brazil and the US, doxxing attacks also resulted in unwelcome product orders being delivered to journalists' homes, including pizzas, sex toys, manure, worms, and other offensive material.

ii. AT THE INTERSECTION OF MISOGYNY AND OTHER FORMS OF DISCRIMINATION

Our research highlights the ways in which other forms of discrimination - such as racism, homophobia and religious bigotry - intersect with sexism and misogyny to worsen and deepen women journalists' experiences of online violence. Within our self-selected sample of respondents, while 64% of white women journalists (compared to 73% of women respondents overall) said that they had experienced online violence, the rates for minorities and additionally marginalised communities of women were significantly higher. For example, 81% of women journalists identifying as Black, 86% identifying as Indigenous, and 88% of Jewish-identifying women journalists, reported experiencing online violence. A similar pattern can be seen when analysing the survey data through a sexual orientation lens: while 72% of heterosexual women indicated they had been targeted in online attacks, the rates of exposure for those identifying as lesbian and bisexual women were much higher - standing at 88% and 85% respectively.

Rate of offline attacks associated with online violence:

20% **All women survey respondents**

53% **Arab women respondents**

Intersectional experience rates of online violence



Highly troubling is the finding that Arab women respondents seem significantly more likely to experience offline attacks, harassment, and abuse associated with online violence than other racial/ethnic groups. Over half (53%) of the women identifying as Arab said they had experienced offline attacks they think were seeded online, compared to 11% for white women and 20% overall.

This pattern of heightened online violence risks faced by women journalists at the intersection of multiple forms of discrimination was confirmed through our interviews, with at least 23 interviewees from around the world highlighting the functions of racist, anti-Semitic, xenophobic, bigoted and sectarian abuse intertwined with misogyny.

Racialised and religiously bigoted misogynistic abuse hurled at women journalists online by the far-right is a global phenomenon and many of our interviewees linked this problem to elected officials creating the enabling environment for this pattern to become entrenched. The broader role of 'patriotic trolling' and the weaponisation of social media platforms to advance far-right ideologies has also been explored by other researchers.

Guardian US technology reporter Julia Carrie Wong, who covers disinformation and the far-right, has been the subject of multiple disinformation-laced attacks orchestrated by far-right groups which abused her with racist and anti-Semitic slurs. In an early episode, she "...caught the attention of the Daily Stormer back when [it] was on the surface web and had a pretty significant audience," she said. "That was my first experience with serious online hate because I am Chinese and Jewish. That became a fixation for the writers and commenters at the Daily Stormer who then proceeded to Photoshop me with horns and a Jewish star, and discussed my racial makeup with quite intense and disgusting fervour."

Racial vilification and structural racism

Women journalists from diverse backgrounds described additional layers of attack based on their attributed race and ethnicity. These factors made them "low hanging fruit" or "easy targets". Two Black women journalists we interviewed in the UK palpably demonstrated the structural racism that is amplified through online violence. They both said that being called the 'N-word' is normalised. Rianna Croxford, an award-winning BBC Investigations correspondent, described the repetitive pattern of racist abuse: "It's not the first time somebody called me the N-word. It won't be the last time. I've had racial abuse... 'monkey', mocking my appearance in different ways."

In South Africa, *Daily Maverick* associate editor Ferial Haffajee was maligned and vilified online in an orchestrated campaign that began in 2016 in response to a corruption investigation she undertook when she was Editor at Large for the local edition of Huffington Post. The attacks on Haffajee leveraged disinformation tactics that included accusing her of being a journalist working for "White Monopoly Capital". This attempt to discredit her was particularly potent in the South African context. Haffajee describes "raw racism and misogyny" including taunts that she said she had never experienced before, such as "Go back to India, this is not your country."

Sectarianism, anti-Semitism, and other forms of religious bigotry

In Northern Ireland, online violence against women journalists has taken a sectarian turn. Sunday World crime reporter **Patricia Devlin** has been reporting on families affected by the ongoing criminal activities of paramilitary groups for 14 years. She has recently watched "usual levels" of misogynistic abuse become more sectarian and terrifying:

> *“ In October 2019, I received a message to my personal Facebook account, and it said ‘Don’t go near your granny’s house in Maghera, Tricia, or you’ll watch your newborn get raped’. And it was signed off in the name of a neo-Nazi group called Combat 18, which in the past has had links to loyalist paramilitaries in Northern Ireland. So it was absolutely traumatising. I have never felt disgust and fear like it. I went to the police. That was the turning point for me. I said, I can no longer brush it off. It’s getting worse. My children are being threatened. ”*

Online violence based on religious and associated cultural factors amplifies the risks to women journalists who are targeted in such ways, especially in countries where perceptions of immorality or blasphemy can trigger backlash, and even offline attacks.

In Pakistan, freelance journalist Youssra Jabeen was accused of blasphemy and ‘defaming Islam’ when she wrote about the fact that female heroes are missing from school textbooks. In the aftermath, she was asked to delete the story as “these accusations can get you traced, killed or kidnapped... It could have affected my workplace and everything,” she said. In Lebanon, TV presenter Dima Sadek told us that being a Shi’ite and publicly criticizing the Shi’ite Hezbollah has made her the victim of many smear campaigns, with trending hashtags such as “ ديما الواطية ” or (“Dima the vile”).

Three of our interviewees received anti-Semitic abuse. And Crister Ohlsson, head of security at Sweden’s Bonnier News - which owns the newspaper *Dagens Nyheter* - said much of the harassment against journalists working for their company is based on anti-Semitic abuse emanating from the racist right: “Our Editor-in-Chief is Jewish, the Bonnier family [the publisher] is Jewish. So that’s one dimension of harassment from the extreme right-wing.”

Gender identity and sexual orientation

In many countries internationally, gender equality is not mandated. In others, commitments to gender equality are being wound back. There are also examples of governments condemning LGBTQ identities. These approaches further intensify the digital risks facing many women journalists.

Sabahat Zakariya, a freelance journalist, points to the intersectional abuse in Pakistan, where sexism and misogyny blend with homophobic vitriol (and where the word 'khusra' is used as a slur for 'trans'): "One thing I get a lot is [abuse] about my short hair. They write comments like, 'oh, you're khusri', or something either homophobic or transphobic. You know they try to find a way to get to you or something that will make you feel insecure about yourself."

Our survey respondents and interviewees overwhelmingly identified their online attackers as men, but there is an emerging pattern of harassment and abuse coming from women in some contexts - especially with regard to reporting on transgender issues. CNN International reporter **Tara John** experienced a three-day wave of abuse after publishing what she and her editor Inga Thordar say was a balanced story on trans rights focused on Scotland's Gender Recognition Act laws. The harassment became so intense that John felt afraid to leave her home:

> ***"The moment the piece went out, it exploded. And this was around the start of the lockdown... I'm not really too fussed when people start trolling me... but this time it felt very personal. I had, for about three days, people scrutinising my tweets. One group threatened to sue me on Twitter. Others told me to f**k off, many encouraged pile-ons and others called me a men's rights activist, a moron, and it just did not stop... Perhaps the intensity of it was, for me as a woman, that these were also women."***

iii. AT THE INTERSECTION OF DISINFORMATION AND MISOGYNY

Disinformation is now inextricably bound up with online violence against women journalists: disinformation methods are co-opted in misogynistic attacks; and reporting on disinformation and associated themes (e.g., digital conspiracy communities; far-right networks) is a lightning rod for pile-on attacks. This helps explain the high number of women journalists responding to our survey (41%) and participating in our interviews who have experienced online violence in the context of what they perceive to be coordinated disinformation campaigns.

Most of the women we interviewed for this research experienced disinformation-based attacks designed to smear their personal and professional reputations, hold them up to ridicule, and expose them to increased offline risk, thereby both humiliating them and potentially chilling their accountability journalism, while simultaneously undermining trust in their reporting. A significant number of the journalists we interviewed had been "slut-shamed" by attackers seeking to cast doubt on the morals of women working in culturally and socially conservative societies. In particular, journalists based in Latin America, Africa, Asia and the Arab States were targeted in this manner - from the use of deepfakes to false narratives about extramarital affairs, and sexual orientation.

Reporting on disinformation is also increasingly recognised as a trigger for orchestrated, cross-platform online attacks - it was a beat identified by 16% of the women we surveyed as one likely to lead to an increase in harassment and abuse. Among them are those reporting on disinformation, far-right extremism, conspiracy and disinformation networks in Brazil, South Africa, the UK and the US.

Investigative journalist **Brandy Zadrozny** covers disinformation for NBC News and MSNBC in the US. Her reporting has focused on far-right networks, the QAnon conspiracy, and the anti-vaccination movement. She "loves" her job, she said, but the impacts of online violence have led her to devise an exit strategy, and she does not plan to be covering disinformation in five years' time:

> ***"I'm not in a war zone, I'm behind a computer. But the effect of online harassment and bad faith journalism from disinformation agents... is meant to silence us, meant to stop our reporting, meant to scare us. I think it does all of those things... And that takes a real toll on you."***

Nationalism, foreigners and spies

Following publication of her BBC investigation into Nigeria's cough syrup cartels, award-winning Nigerian journalist Ruona Meyer was targeted in a campaign of extreme online harassment which lasted almost a year. Her marriage to a German citizen and her association with the BBC led to her being accused of being a 'foreign agent' by anonymous trolls. "I'm also somebody not to be trusted because I'm not patriotic. Why? Because I have...'sexually transmitted citizenship'." Trolls - linked, she suspects, to those the investigation exposed - created a fake blog in Meyer's name, posted homophobic content on it and tagged the BBC in an attempt to elicit a negative response from her UK audience. Separate attacks designed to damage her reputation in Nigeria speak to how the country's gender norms may be used to query women journalists' credibility: "Whenever they want to turn the Nigerian public against me they'll say things like I'm wayward, I married a foreigner, my husband's penis doesn't work... What has that got to do with my work as a journalist?"

In Serbia, women journalists are called 'Albanian' and 'Gypsy' as insults, while in Sri Lanka, the *Daily FT*'s Marianne David - a veteran business journalist - was called a traitor to her country when she spoke out against women being sexually harassed on the street. Maria Ressa is also falsely accused of being a foreigner and a "CIA agent" in The Philippines.

iv. TOP SOURCES AND TRIGGERS ASSOCIATED WITH ONLINE VIOLENCE AGAINST WOMEN JOURNALISTS

The impacts of populist politics, which were associated with misogynistic narratives and the demonisation of journalists and journalism in every region we studied, are clear in our data. This reflects the increase in politically motivated attacks on journalists and independent journalism around the world. The role of political actors - including politicians, government officials, political party representatives, party members, political operatives, and extremists on the political fringe - as instigators and primary perpetrators of online violence against women journalists is an alarming trend confirmed by our survey respondents, interview participants and big data case studies. For example, over a third (37%) of our survey respondents identified political actors as top sources of online attacks against them. Prime targets of such attacks among our interviewees include: Lebanese journalist Ghada Oueiss; Brazilian journalist Patricia Campos Mello; Maria Ressa in the Philippines; Kimberly Halkett, April Ryan, Julia Carrie Wong, and Brandy Zadrozny in the US;

37%

of our women survey respondents identified political actors as top sources of online attacks against them

Carole Cadwalladr, Rianna Croxford, Nadine White, an anonymous political journalist, and BBC Persian journalists in the UK; and Pauli Van Wyk and Ferial Haffajee in South Africa.

In late 2018, **Patricia Campos Mello** started publishing a series of articles investigating disinformation allegations in the context of the Brazilian national elections. The orchestrated attacks against her involved significant disinformation - including that she traded sexual favours for information. The internet became littered with falsified pornographic images of her, deep fake videos, viral memes, and rape threats:

> ***"The type of violence online was very aggressive, it was more false news. There were thousands of memes of a naked woman, or in bra and panties, with my face on it...thousands of memes with different [sexual] positions, movies, videos of me as a 'prostitute journalist', and threats such as 'you should be raped'."***

In response, Campos Mello sued Brazilian President Jair Bolsonaro and his Congressman son for moral damages, accusing them of stating or implying that she offered sex in exchange for exclusive information. She won. A Brazilian court found against the President and his son in early 2021.

While the UK government has recently announced a new framework to better train and empower industry leaders and police forces to address online violence against journalists (among other forms of attack), two Black journalists we spoke to had experienced severe online harassment after their journalism was discredited by the UK's Equalities Minister. **Nadine White** was a reporter at HuffPost UK when the Minister shared screengrabs of the journalist's emails via Twitter, described her as "creepy and bizarre," and said her conduct provided "...sad insight into how some journalists operate." A flood of abuse followed - via email, phone and social media. In response, White and her editor called for a public retraction, and then sent a legal letter to the Minister, asking for an apology:

> ***“ As a Black woman journalist, coming into this industry as the vast minority within this white-dominated elite base, it’s daunting on so many levels just to get up in the morning each day and do what I do, much less to know that you’ve been targeted by a minister and that’s been effectively sanctioned by the people that run this country, by the powers that run this country. ”***

The Council of Europe issued an alert about the incident in response to notifications from several press freedom organisations, but the UK Cabinet office dismissed a formal complaint from HuffPost. The UK replied to the COE alert in March 2021, stating:

The Government agrees that the harassment of journalists and those working in the media is unacceptable and that is why the United Kingdom has established the National Committee for the Safety of Journalists [and]... a National Action Plan for the Safety of Journalists, which looks at measures to address the abuse experienced. The Government notes that these tweets were not issued from an official government Twitter account, but instead from a personal Twitter account. The Minister is personally responsible for deciding how to act and conduct herself, and for justifying her own actions.

Our research confirmed that online violence against women journalists comes from State officials, and is increasingly associated with legal harassment. In Pakistan, founder of the Digital Rights Foundation, Nighat Dad, said: “The attacks are made by people declaring affiliation with the ruling party, and in the coordinated campaigns women journalists are referred to as peddlers of ‘fake news’, enemies of the people and accused of taking bribes. Some journalists [have] shared that after official harassment, their social media accounts are bombarded with gendered slurs and abuse by accounts displaying the ruling party’s flag or the Prime Minister’s picture on their accounts.”

Our participants also revealed an alarming trend involving male journalists from competing news organisations and fringe blogs, especially those from the edges of the political spectrum, ‘dogpiling’ specific women journalists. Eight percent (8%) of our survey respondents identified “staff of rival news organisations’’ among their regular abusers. This

was a particularly noticeable pattern in the US and the UK, with multiple interviewees being subjected to gaslighting attacks from male journalists and commentators featuring pernicious sexism and misogyny.

This behaviour is identifiable as part of a pattern whereby certain influential male journalists trigger or amplify abuse by signalling (e.g., through sharing a tweet from the targeted journalist with critical annotations) to their followers to join the attack on a woman journalist. Niche sites and partisan news media outlets then respond to these cues through blogs, columns or TV segments - both amplifying the abuse and delivering new participants to the 'pile-on' brigade. In some cases, the attack is then "laundered" or legitimised by mainstream media actors. This type of coordinated online assault on a woman journalist sometimes includes pressuring her employer to sack her. *New York Times* technology reporter **Taylor Lorenz** is a regular target of such campaigns across multiple platforms:

"My frustration and anger with all of this is I actually don't even care about death threats anymore. I've had so many people tell me they're going to come rape me and murder me. I'm kind of immune to them by now. But what I really care about is the reputational harm... Their ultimate goal is to make you seem difficult in some way... They wrote a whole story where it implies that I rape children. It's insane. But smears like that hurt my credibility in the eyes of the public... And let's also talk about how the right-wing media amplifies all of it... you would not believe how famous they've made me in that ecosystem."

Thematic reporting triggers

As indicated, gender was the story theme most frequently associated with heightened harassment according to our research participants, highlighting the primary function of misogyny in online violence. Lightning rod issues in the category of gender identified by survey and interview participants include: feminism, domestic violence, sexual assaults, femicide, reproductive rights (especially abortion), and transgender issues. The second most likely theme to be met by online harassment and abuse was 'politics and elections', underscoring the role of political attacks on the press in exacerbating journalism safety threats.

In Sri Lanka, an anonymous interviewee described being attacked on Facebook - through her inbox as well as the pages of political parties - in the context of her posting on political issues:

> ***“ I would be getting these abusive messages, like 100 messages a day, to my filtered inbox…threats all sorts of threats… then other admins reposted it…and sent me messages, voice recordings saying that I would like to do this and this to you, I would rape you, kill you, throw acid at you, kidnap you - beware when you are taking a bus…and that has happened a few times. ”***

Aima Khosa, a news editor at Pakistan’s *Friday Times*, became the target of prolific abuse when she reported about the kidnapping of a young Pashtun activist who had made a speech critical of authorities in 2019. In response, social media users accused her of prostituting herself: “… they started calling me a whore, and my mother a whore, too. Some said ‘your mother is sponsoring you as a whore’, ‘what’s your hourly rate?’… These created long threads with groups of people, men, talking to each other, enjoying a foul conversation about me, against me, attacking me, accusing me … One after another.”

One journalist in the UK, who chose to remain anonymous, was abused in the context of national political party events she was covering while also coming under sustained attack online. In response, her employer hired bodyguards to protect her. By contrast with most other UK journalists we interviewed, she stated: “…the far left has been the worst for me by a million miles…the most alarming bit to me is about how legitimate and professional people have started using attacks on journalists as part of their campaigns. That’s much more alarming to me than the sort of mob stuff either online or in real life, which is horrible.” The journalist still regularly receives messages telling her to “f**k off [eds’ asterisks] and die.”

v. WHEN ONLINE VIOLENCE SPILLS OFFLINE

A pattern emerged from this research, connecting online violence campaigns and offline attacks, highlighting the escalating threat levels faced by women journalists globally. Aside from inflicting very real psychological injury, targeted online violence poses increasing physical safety risks, especially in the context of campaigns against women journalists that trigger mob reactions. One-fifth (20%) of the women journalists we surveyed reported experiencing abuse and attacks in the physical world that they believe were seeded online. This finding is particularly disturbing given the emerging correlation between online attacks and the actual murder of journalists with impunity.

Additionally, the fear and anxiety connected to the threat of physical attack is also relevant. Seventeen percent (17%) of our survey respondents indicated that they had felt physically unsafe as a result of online violence, while 4% reported missing work due to the potential for the online attacks they had experienced to morph into physical attacks. Many of our interviewees also discussed feeling compelled to adapt their offline movements and increase their physical security in response to serious threats made online. Some even relocated multiple times, with significant impacts on their families.

In Northern Ireland, *Sunday World* reporter Patricia Devlin has received online death threats in parallel with offline threats - including having her name graffitied on walls in Belfast alongside the crosshairs of a gun. The police also warned her about multiple death threats in 2020. However, 17 months after reporting the aforementioned rape threat against her newborn son to the police there has been no progress on the matter: "It was traced to an individual but he's never been arrested or questioned, and he's actually fled the UK now," she said. "And I have a complaint with the police ombudsman about that case because I feel they inadequately investigated it and their actions led to this person absconding. [He] has a serious history of violence and links to a paramilitary group here in Northern Ireland. He's carried out attacks. He's also involved in the neo-Nazi movement in Scotland - a really dangerous individual."

Also in the UK, BBC Disinformation Reporter Marianna Spring was stalked in connection with her COVID-19 conspiracy debunking reports and social media commentary. A message was left for her on a notice board outside the train station she uses to commute to work. "So when I'm walking to and from work, it's me that has to be hyper-aware, me that can't listen to music or podcasts when I'm walking anymore. It's me that kind of has to forfeit certain freedoms as a consequence of these people being horrible," she said.

20%

of all women survey respondents experienced abuse and attacks in the physical world that they believe were seeded online

17%

felt physically unsafe as a result of online violence

Sri Lankan author and former journalist Sharmila Seyyid was targeted when she suggested in an interview with the BBC that the legalisation of sex work could provide workers some protections. She was then subjected to a campaign of harassment by self-identified Islamic fundamentalists. Seyyid has been 'raped and killed online', cyber stalked, and had her identity stolen. The online campaign against her climaxed in a group circulating false stories that she was dead, along with Photoshopped images of her apparently mutilated body. The ongoing threats made against her, the hate campaigns and propaganda have had significant psychological, physical and professional impacts on her, she said. Her case illustrates the ways in which online violence can be paralleled by, or trigger, offline attacks: she has been mocked in public - even while shopping; the English academy that she and her sister ran was vandalised; and her parents have had to relocate several times. Eventually, in response to this onslaught, she gave up her journalism career and fled the country to India for three years with her infant son.

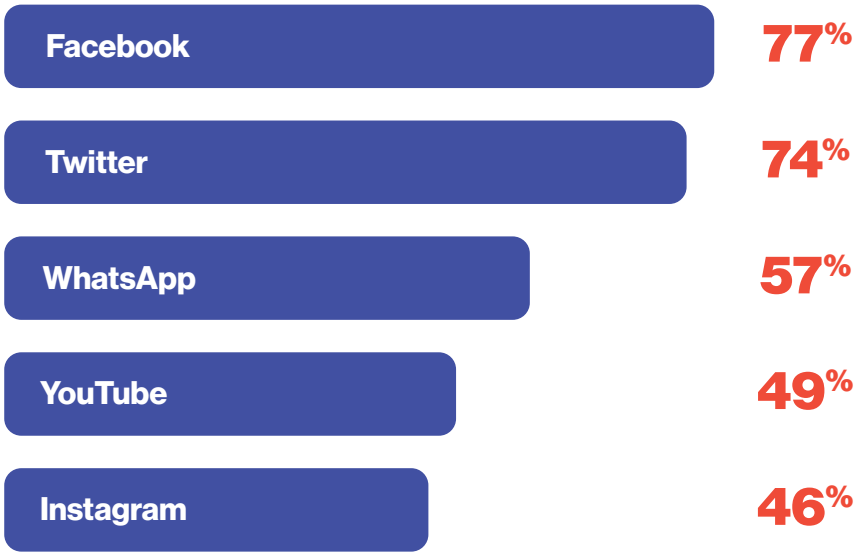
For The Grio's White House correspondent April Ryan in the US, the offline impact of the online violence she experienced included people coming to her home, or waiting for her outside the White House to abuse her when she finished work. As a result, she cannot stop looking over her shoulder: "I'm in therapy and they say it is trauma, not only from Donald Trump, but the minions, always having to...make sure someone's not coming after you," she said. "Will my life ever be the same? No." Two of our other US-based interviewees were physically attacked - punched in the face while walking in the street or reporting on far-right rallies - in the course of their work. Both were investigating radicalised supporters of former President Donald Trump.

It is no wonder, then, that so many of the women we interviewed around the world had resorted to changing their patterns of movement and even going into hiding in response to online violence. Many also increased their physical security, highlighting their sense of vulnerability and their awareness of the potential offline consequences of digital attacks.

vi. THE PLATFORMS: VECTORS OF ONLINE VIOLENCE AGAINST WOMEN JOURNALISTS

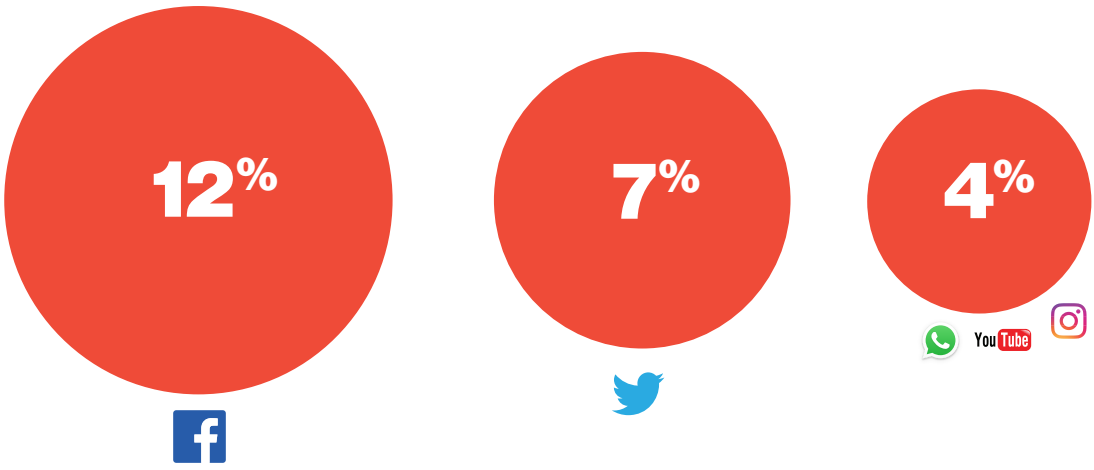
The role of internet communications companies as vectors for online attacks against women journalists cannot be underestimated in an era of digital journalism, networked disinformation, toxic online communities, and populist political actors who weaponise social media and leverage misogyny as tools to demonise journalists. According to our survey, the top five platforms or apps most frequently used by the women survey

Top 5 platforms most frequently used by women survey respondents[13]



Very unsafe high-use platforms

As indicated by our survey respondents who identified as women.



[13] WhatsApp and Instagram are both owned by Facebook.

respondents in the course of their work were ranked like this: Facebook (77%); Twitter (74%); WhatsApp (57%); YouTube (49%); and Instagram (46%). But Facebook was disproportionately identified as the platform to which respondents most frequently reported online attacks (39%), with Twitter attracting complaints at the rate of 26%. Facebook was also identified as the least safe of these high-use platforms, with 12% rating it "very unsafe" - almost double the number (7%) who rated Twitter "very unsafe". WhatsApp, YouTube and Instagram were identified as "very unsafe" by 4% percent of respondents.

Considering the role of Facebook and Twitter as major enablers of online attacks against women journalists, the fact that just over a third of our survey respondents had formally reported the problem to Facebook, and just over a quarter to Twitter, probably reflects both a sense of futility frequently associated with such efforts and the general reluctance of the women surveyed to escalate these issues externally. Of particular concern is the fact that even though social media platforms are expected to comply with international human rights standards and have policies against online abuse, when women journalists report online attacks, the platforms often fail to adequately enforce their own rules.

Our interviewees, across all countries examined in this study, voiced very significant dissatisfaction with the social media companies' responses to abusive content and accounts that they reported through the platforms' facilities. The most common experience they mentioned was that the platforms frequently ignored or rejected their requests for the deletion of offensive content or accounts. For example, the South African editor Ferial Haffajee said she was frustrated by what she termed "stonewalling" by Twitter's automated reporting system, which did not provide allowances for local context or sensitivities. But Filipino-American journalist Maria Ressa reserved her harshest criticism for Facebook, and others pointed to YouTube as the worst offender. Such concerns underpinned the widespread call among our interviewees for these companies to employ many more human moderators with specialist training in human rights, particularly in the areas of gender equality and press freedom.

In order to force the companies into action, **Susanna Skarrie**, Editor-in-Chief of Swedish magazine *Hem och Hyra*, enlisted help from an external consultant to liaise effectively with Google and Facebook about removal of abusive content and reduction of traffic to websites targeting her, her family, and her colleagues with abuse:

> *“It is way too complicated to deal with this yourself and it requires constant work. At first Google said they would stop the search engine optimisation of these websites but then they slip through anyway and are back online again. And Google does nothing. Every time a new subpage emerges we have to contact Google again.”*

Skarrie’s dealings with Facebook were similar: “We tried to get Facebook to take down these accounts and at first they simply say they are not illegal, although they clearly are. While we did manage to get some of these accounts taken down, they keep reemerging and are used to slander me and my family and other journalists who have investigated them.”

We were also told by interviewees in the Arab States region that the scope and swiftness of platform responses can depend on the amount of attention the journalist/incident is getting, their prominence on the platform (e.g., the number of followers they have), and whether or not international organisations were drawing attention to their cases.

What all this clearly demonstrates is the need for social media companies to stop making reporting or blocking offenders the responsibility of the journalist herself. Instead, the women we interviewed argued that the platforms need to respond proactively to acts of online violence, including gendered violence, and work harder to prevent such expressions and their dissemination, not wait for them to make complaints on a scale deemed necessary to trigger action.

Another significant concern raised by our interviewees was the unevenness with which these companies are detecting and acting on abuse across different languages and countries. While abuse moderation in languages spoken in the companies’ major markets is somewhat better addressed, this is far from being the case in less prominent local languages, including in countries such as Kenya, Nigeria, Sri Lanka, the Philippines and Pakistan. At an intersectional abuse level, interviewees pointed out that being abused in languages such as Persian, Sinhala, Tagalog, Malay, Urdu or Tamil adds to the frustration of trying to report abuse to the platforms which have very limited abuse reporting and moderating capability in these languages, and therefore often do not respond appropriately. This is a point underscored by Pakistani journalist Youssra Jabeen: “There is no point in us reporting anything because we know nothing is going to get done there. They operate in English, so how do you report threats in Urdu?”

In addition to linguistic diversity, the cultural context of online violence is currently not captured by the platforms' algorithms and policy implementation workflows. A particularly pointed example is the experience of the Al Jazeera journalist Ghada Oueiss, who reported to Twitter a fraudulently altered video purporting to show her naked in a jacuzzi, screengrabs from which she said were retweeted 40 thousand times. The video and a series of photos showing her eating a meal with colleagues were stolen from her phone, she alleges, as part of an orchestrated attack designed to discredit her. They were distributed with messages alleging she was an alcoholic, drug-addicted prostitute. Twitter's North American corporate perspective was that the content did not violate their policies. However, being a journalist working in the Arab States, the wide circulation of this stolen and manipulated material put Oueiss at risk of retribution and caused significant reputational damage.

Many of our interviewees and survey respondents said the platforms need to create rapid response units - staffed by multilingual employees with expertise in press freedom and gender-based violence. Others called for the equivalent of a "big red button" that they could hit when they came under sustained attack on a social media platform, or received a serious threat. A series of practical product design solutions proposed by PEN America represents a potentially effective set of responses at a technical level.

It should also be noted that our interviewees' attitudes to the platforms varied. While Facebook was generally condemned by the women journalists we interviewed, some (e.g., Maria Ressa in the Philippines) reported recent improvements in their dealings with Twitter, while others reserved their harshest criticism for Google and its social video channel YouTube.

There is a pattern of victim-blaming associated with online violence against women journalists, which seeks to make the targets responsible for the toxic behaviour of others, paralleling the sort of treatment women still often encounter in the context of sexual harassment, assault or domestic violence. This is reflected in the "mute, block, report, delete" mantra of internet communications companies that are failing to deal with the problem structurally, and it is also evident within law enforcement and judicial contexts, as well as within news organisations.



vii. NEWS ORGANISATIONS STRUGGLING TO RESPOND EFFECTIVELY TO ONLINE VIOLENCE

Despite the high incidence of online attacks identified by the women survey respondents, only a quarter (25%) reported incidents to their employer. This likely indicates that: women journalists, who are frequently structurally disempowered and disadvantaged in newsrooms, remain reluctant to draw attention to the impacts of online violence with their employers; and it is only the more extreme instances that cause women journalists to escalate such safety concerns within their workplaces.

The responses that the women survey participants reported receiving when they did report online violence to their employers were, on the whole, very unhelpful. They illustrate the enduring failures within many news organisations to respond appropriately or effectively to the crisis. The most common response received was that no action was taken; the next most frequently identified response was gender-insensitive advice such as "toughen up" or "grow a thicker skin". A Kenyan journalist who chose to remain anonymous was clarion on this point:

> ***"Having a thick skin does not protect you from a personal attack that leads to your data being shared and someone promising that they will rape you."***

More disturbingly still, some of our research participants said their employers asked them what they did to provoke the attacks, highlighting the perpetuation of victim blaming, along with women's subjugation. This also demonstrates a clear need for more gender-sensitive approaches to managing the problem within news organisations.

The women journalists who participated in our survey identified mental health impacts as the most significant consequence of the online violence exposure that they had experienced (26%). And a substantial number (12%) said they had sought medical or psychological help in response. Another 11% said they had taken time off work to recover, and a number of our interviewees experienced severe psychological injury, including PTSD. Indeed, many of our interviewees broke down while discussing their experiences with us, describing how important it was for them to be listened to.

12%

of women survey respondents sought medical or psychological help

Access to psychological support is therefore crucial, but of the women journalists participating in the survey, only 11% said their employer provided access to a counselling service. In many newsrooms, sometimes even where a protocol for responding to online violence was in place, our research revealed a lack of understanding of the psychological harm caused by such attacks, and a corresponding failure by employers to address the potentially serious mental health consequences for those targeted. Historically, professional cultural barriers have also been significant.

Kristen Neubauer from Thomson Reuters said: "There's this culture and a tradition … in the news industry… a kind of a macho swagger. [When we first launched our peer network], people were reluctant to talk about problems for a variety of reasons: [they were] concerned about how that would reflect on their career and [thought they might be] pulled off stories or assignments; and … there's a stigma [about] mental health and discussing problems. … And so nobody came to talk to us in the beginning."

There is also a false narrative that insists that online violence is not "real", or that it is incomparable to physical violence. In Serbia, journalist Jovana Gligorijević said the lived experience of women journalists exposed to online violence is frequently diminished by those who covered the 1990s conflicts in the former Yugoslavia: "When it comes to taking online threats seriously, that will never happen. I talked to [my colleagues] very openly…to let them know that they need to acknowledge that… I don't feel safe and that [online violence] happens. The answer I got comes down to one sentence and that is: 'You know how we were all war reporters, so to us all this is like nothing. [It's] nonsense compared to what we went through."

A number of larger international news organisations have formal, written policies and procedures in place to deal with online violence, particularly in the US and the UK (e.g., the *New York Times*, *The Guardian* and the BBC). Some smaller news organisations, such as Rappler and Vera Files (Philippines), the Center for Investigative Reporting (Sri Lanka) and the *Premium Times*' Centre for Investigative Journalism (Nigeria) also have effective practices in place to counter gendered online violence despite the absence of formal policies at the time of writing.

However, many of our interviewees - even those working for international news organisations regularly held up as exemplars of good practice - expressed exasperation and a sense of abandonment by their employers when they were in the midst of an online violence storm. This was generally a response to gender-unaware policies, or protocols that had stagnated as a result of a failure to take account of increasing online toxicity, escalating political polarisation and populism, the rise of far-right extremism, and viral disinformation which have taken hold globally since 2016.

One of the most significant grievances expressed by our interviewees was about the victim-blaming and 'speech-policing' in evidence as a response to online violence - even in major international news organisations. Sometimes, this was a product of social media policies requiring journalists not to engage with their attackers, or to avoid expressing any kind of opinion about controversial matters, under the misapprehension that this would prevent trolling.

Many of our interviewees also said they wanted a newsroom "point person" whose job it would be to take over their email and social media accounts in the event of a major online violence attack. They need their accounts monitored by a trusted person to minimise further exposure to abuse and threats, preferably by someone who can also catalogue threats and screenshot abuse in case of future legal action. The "point person" should also contact social media platforms on the journalist's behalf to escalate reports and aid efforts to trace users behind the most serious attacks, they suggested.

One very clear point of agreement across all women journalists experiencing intersectional abuse came back to the culture and leadership in many newsrooms. Sri Lanka's Methmalie Dissanayake, Deputy News Editor at *Ceylon Today*, suggested that women do not often raise these issues publicly, or in the workplace. Fatima Hussein, president of the Washington Baltimore News Guild and reporter for Bloomberg Law, said that "traumas" are routinely absorbed by women facing intersectional discrimination. "Since there is this lack of diversity in editorial ranks and among reporters, even the notion of complaining about ill treatment is frightening because I want to keep doing my job and I don't want something to be taken from me. I don't want my beat to be taken from me."

South African Editor-in-Chief, Branko Brkic (*Daily Maverick*) may provide a model for (mostly male) editors to follow in responding to gendered online violence against their staff. Investigative journalist **Pauli van Wyk** credits Brkic's leadership and support for her survival as a journalist when she faced large scale online attacks in the course of her reporting:

> ***"[He was] incredibly supportive. He really was quite amazing. He shielded me from attacks and he helped me to move around the country [in response to credible threats] as well. And without my editor, I can tell you now, I probably wouldn't be a journalist anymore."***





Maria Ressa

At the core of an online violence storm

Photo credit: Franz Lopez/Rappler

> ***First, I'm attacked for being a journalist, second I'm attacked for being a woman.***
>
> Maria Ressa

Death threats. Rape threats. Doxxing. Racist, sexist, and misogynistic abuse. In text, image and memes. These are just some of the features of the online violence that Maria Ressa, the Filipino-American journalist who founded the Manila-based news site Rappler, has faced daily since the Philippines' 2016 election. Ressa is an emblematic case study in the global scourge of online violence against women journalists. She lives at the core of a very 21st century storm. It is a fusillade of disinformation and attacks - one in which credible journalists are subjected to online violence with impunity; where facts wither and democracies are put under severe strain. The former CNN war correspondent and laureate of the 2021 UNESCO World Press Freedom Prize says none of her experiences in the field prepared her for the massive and destructive campaign of gendered online abuse, threats, and harassment directed at her over the past five years. At one point, in response to an investigative series on State-linked disinformation, she recorded receiving more than 90 hate messages an hour on Facebook.

This collaborative case study[14] represents a forensic analysis of hundreds of thousands of Facebook and Twitter posts between 2016 and 2021 directed at the internationally celebrated digital media pioneer. It highlights the intensity and ferocity of attacks designed not only to vilify a journalism icon, but to discredit journalism itself, and destroy public trust in facts. These attacks also created an enabling environment for Ressa's persecution and prosecution in the Philippines. Now, her life is at risk and she faces decades in jail.

[14] A detailed version of this Participatory Action Research case study (including methodology and dataset descriptions) is available here: https://www.icfj.org/our-work/maria-ressa-big-data-analysis

Ressa is not only attacked for being a journalist. She is attacked for being a woman. For the colour and texture of her skin. For her dual citizenship. And for her sexuality. The attacks against Ressa are enabled by a climate in which President Rodrigo Duterte has publicly condemned Ressa - while also once musing that journalists are not exempt from assassination.

The worst attacks appear to have been orchestrated.

Despite belated and largely ineffective attempts to address hate speech within the social media ecosystem, the technology companies concerned serve as the vectors for these attacks, which in turn create a facilitative environment for State-led legal actions against Ressa. This 'lawfare,' as she refers to it, led to her conviction in mid-2020 on a criminal 'cyberlibel' charge, and it continues to escalate. Issued with 10 arrest warrants in less than two years, and detained twice in the space of six weeks, Ressa is currently fighting nine separate cases. If she is convicted on all charges, she could spend the rest of her life in jail. **Caoilfhionn Gallagher QC**, the co-lead of Ressa's international legal team, states that for having published journalism, her client now faces a barrage of baseless lawsuits that seek to criminalise her work and expose her to a century in prison:

> ***"State authorities thus both directly attack Maria, and also create an enabling environment that facilitates and fuels abuse from others. In turn, online abuse emboldens the authorities in their persecution of her. In my view, there is a symbiotic relationship between the abuse Maria experiences online and the progress of the legal harassment offline."***

This case study combines big data analysis - using Natural Language Processing (NLP) techniques (merging linguistics, computer science and Artificial Intelligence) and network analysis - with deepdive interviews and extensive desk research. Our analysis of the online violence Ressa experiences helps us understand how and why she is targeted, and how the threats spread. It also gives us insights into the role of State actors in manufacturing consensus by fuelling the behaviour of the social media users who target her. “If I wanted to see what the government was going to do, I only needed to look at social media because the attacks to arrest me and shut down Rappler were seeded as meta-narratives in 2017,” Ressa said. “And now here we are.” The function of the social media companies in facilitating the abuse is also spotlighted by Ressa: “The only way it will stop is when the platforms are held to account, because they allow it... They have enabled these attacks; they should not be allowing this to happen.”

None of her experiences in the field prepared her for the massive and destructive campaign of gendered online abuse, threats, and harassment directed at her over the past five years.

12 KEY FINDINGS

from our examination of online violence against Ressa over a five-year period (2016 to 2021)

Attacks designed to discredit Ressa's journalism:

// **Almost 60% of the attacks on Ressa extracted from Facebook and Twitter for analysis were designed to undermine her professional credibility and public trust in her journalism.**

// **Credibility or reputation-based attacks frequently deployed disinformation tactics and abuse conflating Ressa and her journalism with "fake news".**

Attacks designed to undermine Ressa's personal dignity:

// **Over 40% of the attacks in the combined datasets targeted Ressa at the personal level - often viscerally.**

// **The most prevalent type (34%) of personal abuse against Ressa is sexist, misogynistic and explicit. Racist and homophobic abuse accounted for 5% of personal abuse.**

// **14% of all of the abuse studied is sexist, misogynistic and explicit.**

Networked and technically advanced abuse:

// There is direct evidence that the online violence targeting Ressa has offline consequences. It has created an enabling environment for her persecution, prosecution and conviction. It also subjects her to very real physical danger.

// The use of abusive memes and manipulated images, which 'fly under the radar' of detection, is commonplace.

// There is evidence that the worst of the attacks on Ressa involve coordination or orchestration - often associated with disinformation campaigns led by State actors.

Sources, triggers and impacts of online violence against Ressa:

// Much of the abuse and threats are fuelled by statements and messages from political leaders and partisan bloggers/social media influencers who demonise Ressa and Rappler as unpatriotic criminals. It is also amplified by pro-government news media.

// Lightning rods for attacks include Rappler's investigative journalism; Ressa's reporting and commentary on State-linked disinformation; Ressa's high-profile media appearances; her industry accolades; and her court appearances.

The role of the platforms:

// Facebook is the main vector for the online violence Ressa faces. It is also the most used social media site in the Philippines.

// Both Facebook and Twitter have promised to address the attacks on Ressa but during the period under review, Facebook failed "woefully", she said, to effectively stem the tide of hate against her. However, she now feels "significantly safer" on Twitter.

A climate of impunity

The Philippines is routinely ranked as the most dangerous country in Southeast Asia for journalists, and UNESCO's Observatory of Killed Journalists identifies 109 fatalities since 2006, with 83% of cases currently unresolved. Targeted online violence attacks against journalists like Maria Ressa thus need to be examined in the light of the impunity situation. In 2009, the country was the site of the deadliest attack on journalists ever recorded by the Committee to Protect Journalists (CPJ) - the Maguindanao massacre, which killed 30 journalists and media workers[15] in an "orgy of political violence." Immediately before President Rodrigo Duterte took office in 2016, he issued threats against journalists and a year later, he criticised Ressa in a State of the Nation Address. Since 2017, political influencers have been seeding meta-narratives on social media, painting Ressa as a criminal and calling for her arrest. They were also calling for her to be sexually assaulted, killed and even "raped repeatedly to death."

In a country like the Philippines, where extrajudicial killings continue to be condemned by UN actors, the potential is high for online violence against women journalists to reap deadly results. As Rappler's Executive Editor Glenda Gloria noted: "We never doubted that those online threats would translate to physical threats. That's why we doubled not just the security of Maria, but of the newsroom, because a lot of the online threats against activists turned into reality. There was this female activist who was first blasted online and shot while on her way home. It's real. Especially against women."

UNESCO OBSERVATORY OF KILLED JOURNALISTS

83%

of cases currently unresolved in the Philippines

109

journalists killed in the Philippines since 2006

i. "SHARK TANKS" FULL OF DATA

In this case study, we examine nearly 400,000 tweets directed at Maria Ressa during a 13-month period from December 2019 to February 2021, along with more than 56,000 posts and comments published on Facebook between 2016 and 2021.

The tweets were gathered and analysed using Natural Language Processing (NLP) techniques. From the large dataset, we extracted a sample of 1,128 tweets demonstrating highly explicit abuse - predominantly expressed in English - for detailed analysis. In parallel, data analysts from the digital research firm Graphika conducted a network analysis on a subset of tweets from a spike of abuse in June 2020 that coincided with Ressa's conviction on the first cyberlibel charge. The aim of this analytical process was to map the types and methods of attack, along with the trajectory of the abuse and the interconnectedness of those attacking Ressa.

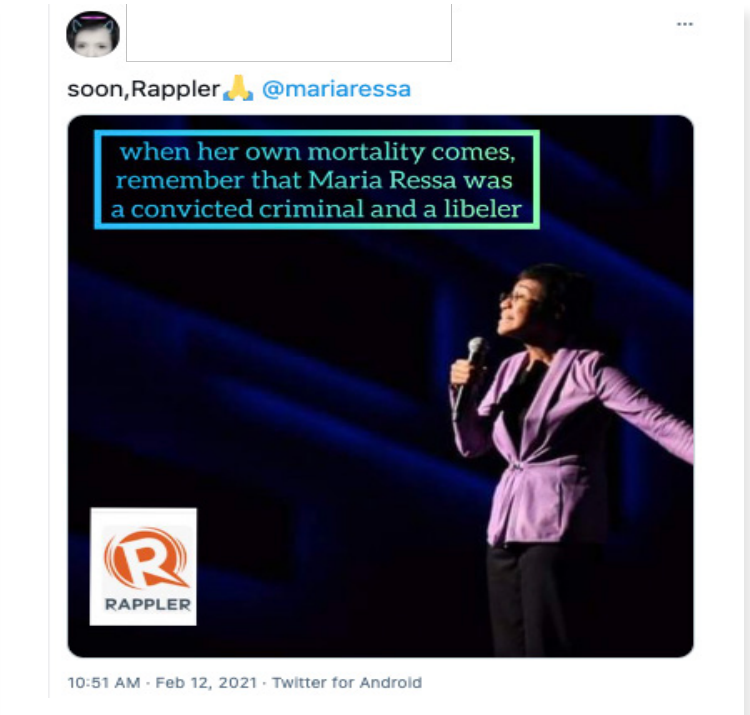
[15] UNESCO officially records 30 journalists murdered during this massacre: https://en.unesco.org/themes/safety-journalists/observatory/country/223790

The Facebook data represents over 9,400 public comments in response to Ressa's posts that were gathered from her professional Facebook page with her explicit permission, and over 47,000 other public Facebook posts mentioning Ressa extracted from a massive database that Rappler maintains, which they call 'Sharktank.' According to Ressa, 'Shartank' maps the information ecosystem of the Philippines on Facebook (which is synonymous with the internet in the country, reaching 96% penetration in 2021). As of January 2021, the database had captured 471,364,939 public posts and 444,788,994 public comments made by 4,176,326 users, 68,000 public pages and 26,000 public groups on Facebook. The same NLP techniques were applied to both sets of Facebook data with the assistance of Rappler's research team.

The NLP analysis is based on a 'high accuracy detection' model and it is largely restricted to English language posts, or those which blend English with Tagalog (the other national language of the Philippines). Consequently, the samples of online abuse extracted using this method are considered to be severely underreported, capturing only around 50% of all English-language abusive messages present in the target's social media stream, according to previous studies.

It is important to note that this analysis excludes the most brutal online violence Ressa has experienced, which she says came via Facebook Messenger. Such content is not only harder to detect and save for big data analysis, it is also more difficult to report because automated abuse reporting systems have very limited multimedia capability and they do not accommodate nuance. Similarly, public abuse is often subtle and potentially designed to 'slip under the radar.'

They were also calling for her to be sexually assaulted, killed and even "raped repeatedly to death".



A death threat sent to Maria Ressa via Twitter on February 12, 2021. The threat was finally removed by Twitter in mid March 2021.

A classic example of this is a death threat sent to Ressa on Twitter on 12 February 2021. This is subtle partly because it carries text in an image which is not easily processed by automated tools, and partly because the message itself contains no abusive terminology, though the underlying meaning is clear. This message was finally deleted in mid March 2021 after being reported to the platform by the lead author and others.

> ***Every time a complaint reaches the court, every time a statement is made in support of Maria, there's a troll army that really is commanded to respond.***
>
> Glenda Gloria
> Rappler's Executive Editor

ii. METHODS, THEMES AND TROPES DEPLOYED BY THE ATTACKERS

While the vast bulk of online violence that Ressa experiences occurs via Facebook, the most common themes and methods of attack were relatively consistent across both platforms. We triangulated this data with online research and in-depth interviews with Ressa and her colleagues. An examination of the combined datasets surfaced the following findings:

Dominant themes and tropes

// Abuse damaging to Ressa's professional reputation or credibility includes disinformation designed to discredit her as a journalist and erode trust in her journalism. It features calls for her to be charged, tried, raped, imprisoned etc. for her work. This abuse frequently involves false claims that she is a purveyor of "fake news" and includes the pernicious hashtag #presstitute.

// Misogynistic, sexist and explicit abuse includes abuse targeting Ressa's physical appearance (emphasising her skin condition) and manipulated photographs depicting her head associated with male genitalia.

// Racist abuse and memes constitute 3% of the personal abuse Ressa receives.

// Homophobic slurs designed to question her sexuality and increase her vulnerability were determined to represent 2% of personal attacks.

// Threats of physical violence, including death threats embedded in images, and threats of sexual violence (e.g., being "publicly raped to death") were associated with the worst attacks.

Triggers for attacks

The attacks spike in association with:

// Rappler's investigative journalism focused on the 'drug war', and the erosion of democracy in the Philippines.

// Ressa's reporting and commentary on disinformation and the government.

// Ressa's high-profile media appearances.

// Ressa's international awards and civil society statements of support.

// Ressa's court appearances.

Typical methods of attack

// Key significant attacks appear to be orchestrated (with the detectable use of fake and bot accounts), and on occasion this has led Facebook to remove networks of accounts identified as participating in what they call 'coordinated inauthentic behavior'. However, the company's response to the attacks on Ressa has been inconsistent and "woefully inadequate," in her words.

// Hashtags designed to encourage swarms of attackers and fuel 'patriotic trolling' are frequently used, and sometimes include abuse within them e.g., #ArrestMariaRessa.

// Memes and manipulated images are deployed to increase engagement with the attacks on Ressa and avoid automated abuse detection tools.

// Doxxing is used to motivate Ressa's online attackers to attack her offline as well.

Spikes and triggers

The largest surge of attacks we identified against Ressa on her Facebook page occured in October 2016, when Rappler published a three-part investigative series into State-linked disinformation networks - two of which were written by Ressa herself. The Facebook data covers a five-year timespan and clearly demonstrates other attack spikes connected to Rappler's critical coverage of the 'drug war' and the extrajudicial killings associated with it, along with international media attention (e.g., associated with the multi-award-winning film A Thousand Cuts), her high-profile industry awards, and her arrests, detentions, and trials are also associated with increased attack spikes in the Facebook data.

An examination of the Twitter data detailing frequency of abuse reveals three major spikes, each of which have more than 50 tweets per week identifiable as highly explicit abuse (largely in English or hybrid English-Tagalog). The largest peak was in early May 2020, and it was triggered by an interview with ABC Australia in which Ressa misspoke, and the error was used to attack her credibility. The next biggest spike came in June 2020, when a Manila court delivered a guilty verdict against Ressa in a criminal cyberlibel charge prosecuted by the government. Her chastisement of the government due to the Philippines being recorded as having the highest proportion of COVID-19 cases in Asia in August 2020 represented the third-biggest spike.

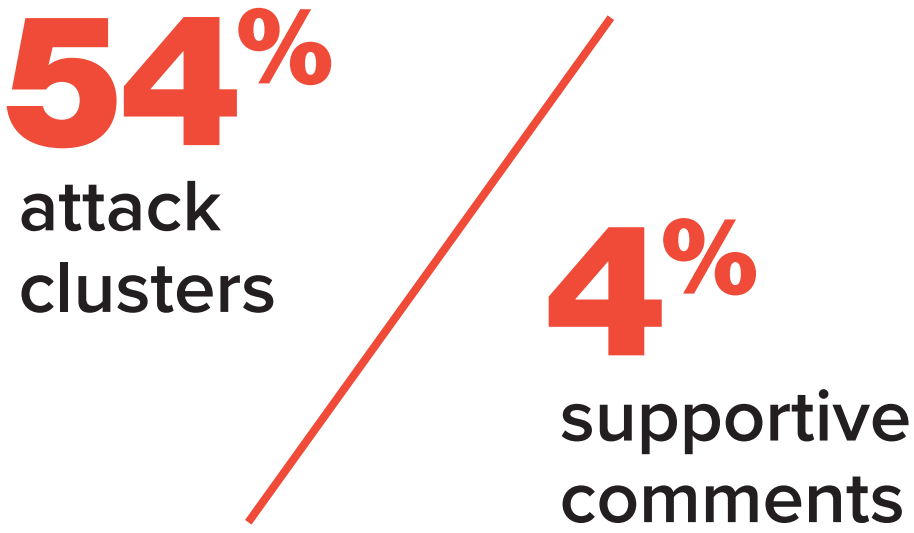


The 100 most prevalent abuse terms in the analysed twitter data. Term size reflects frequency of occurrence.

iii. SCROTUMS, MONKEYS, SWARMS AND LIES

Attacks against Ressa and Rappler dominated the comments on her professional Facebook page, which she established in 2015. According to Ressa, Facebook had recommended she start the page to help better manage her comments. But the harassment she experienced on the page soon became overwhelming. She has not posted to the page since early 2019, and it now lies effectively dormant.

Of the 9,433 comments from Ressa's professional Facebook page - spanning the period 2015 to 2018 - 54% fall under 'attack clusters', while supportive comments represented only 4% of the data. This means that for every one comment supportive of Ressa, there were about 14 comments attacking her. And a more granular analysis revealed that approximately four of these 14 abusive comments would constitute personal attacks, focusing on her appearance, nationality, gender and sexuality.



Attacks designed to discredit Ressa's journalism and erode public trust in facts

As indicated above, the dominant theme of the online violence waged against Maria Ressa involves damaging her professional credibility, by extension her reportage, and by association Rappler's. This is clearly evident within a smaller subset of highly explicit abusive comments (mostly in English or English-Tagalog) gathered from her professional Facebook page, with nearly half of that abuse falling into the category of attacks on her professional reputation.

Among these comments were disinformation-laced attacks, including accusations that she was a 'fake news' peddler, like these:

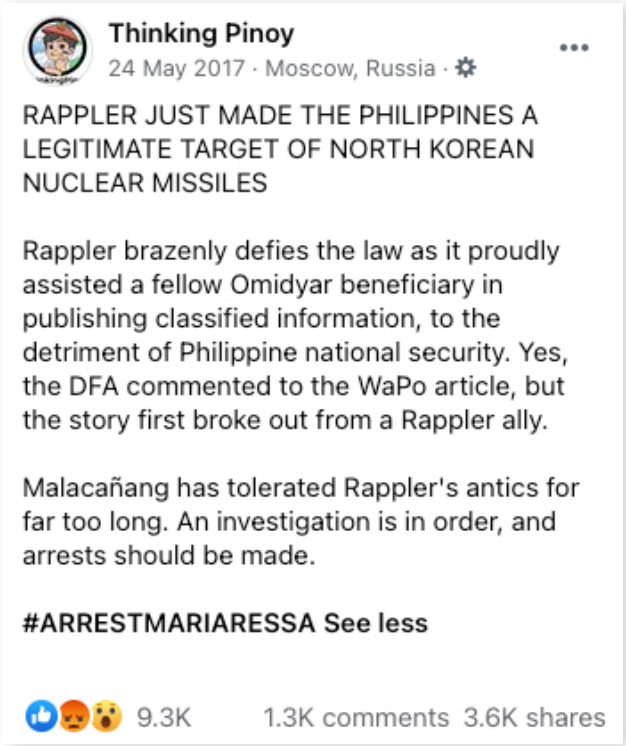
> *"You are the Queen of Fake News f**king [eds**] Bitch (sic)"*

> *"Stop spreading Lies you Piece of SHITS (sic) I wish you Rotten (sic) in Jail ..."*

> *"Maria Ressa get the f**k [eds' asterisks] out of our country Philippines! dont (sic) mislead the people with your fake news ..."*

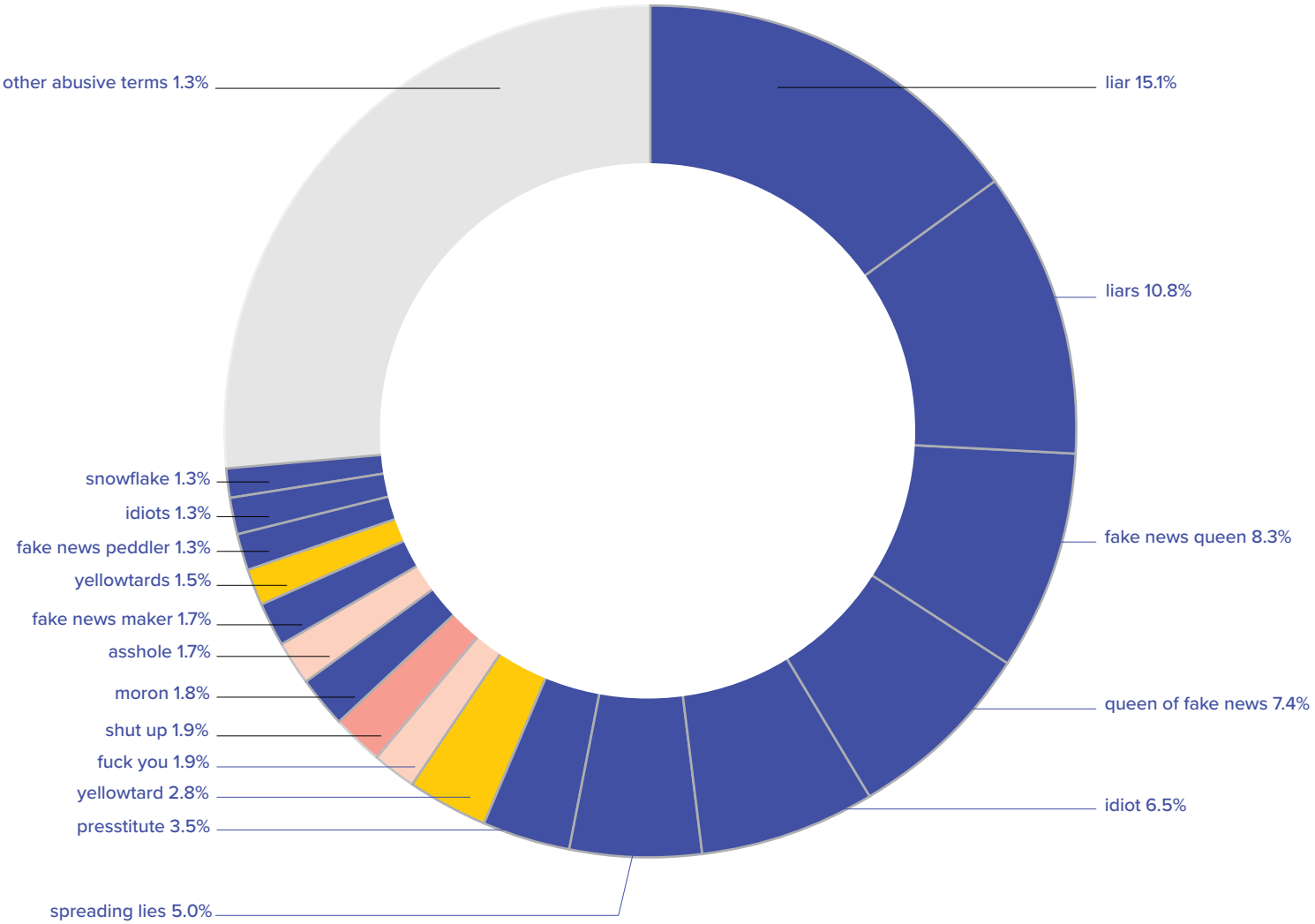
The most frequently used abusive terms in this data subset from Ressa's Facebook page were words designed to ridicule, silence and discredit her while simultaneously undermining public trust in her critical journalism. The top ranked words were "idiot," "shut up," "presstitute," and variations on "liar." Around 20% of the attacks on her credibility were related to disinformation - either equating her with it, or falsely accusing her of peddling it. Ressa has assessed that: "Lies spread faster than facts. And lies laced with anger and hate spread faster and further than facts."

Disinformation narratives deployed against Ressa were also prevalent in the larger multilingual dataset extracted from her professional Facebook page. These include repeats of claims that she is a "liar"; the "Queen of Fake News"; "Bayaran" (a Tagalog term for a corrupt journalist who takes payment for favourable coverage); a "presstitute" ; and a "national security threat" or terrorism supporter, echoing narratives from a prominent Philippine blogger, known as Thinking Pinoy, who tried to get the hashtag #ArrestMariaRessa to trend in May 2017. Two years later Ressa was in fact arrested, and within three years she would be convicted on a criminal cyberlibel charge prosecuted by the State.



A screengrab of a Facebook post published by the. pro-government blogger 'Thinking Pinoy'[16].

[16] In line with academic research ethics protocols, we have obscured the identities of most social media users. In this instance, however, we have not redacted the user's identity as a matter of public interest. 'Thinking Pinoy' is identified in the data and in media coverage linked throughout this report as one of the key pro-government influencers targeting Maria Ressa and Rappler.



This 'donut chart' visualises the frequency breakdown of abusive terms that appear more than once in Facebook posts mentioning Maria Ressa contained within the large 'Sharktank' dataset ("other abusive terms" includes those which appear between one and 15 times in the data). Personal attacks consist of sexist, misogynistic and explicit sexual terms (light pink); and other kinds of personal insult (dark pink). Political attacks (yellow) use terminology associated with (real or imagined) political affiliation. Credibility attacks are coloured in blue.

Online violence against Ressa undoubtedly impacts on decisions in the progress of the cases and charges she faces according to her international lawyer, **Caoilfhionn Gallagher**, who said that when the administration is making decisions in that environment, the fact that Maria is a hate figure online is enabling those decisions to be taken. Referring to the decision to charge Ressa with a third count of criminal cyberlibel in January 2021, Gallagher said:

> ***“It’s impossible not to be aware of the fact that Maria was the subject of a very large amount of viral abuse throughout that time, when the decision was being made. There was a coincidence in time between when the prosecutor was making that decision about the third charge, and a spike in abuse linked to the arrest warrant for Maria in respect of the second cyberlibel charge.”***

In Ressa’s view, the State “propaganda machine” has accomplished its goal: “They pounded opponents and journalists to silence to create a bandwagon effect for seeded meta-narratives of bias, incompetence, criminality, and corruption to be leveled against them.”

According to Rappler’s Managing Editor Glenda Gloria, Ressa transformed into a “warrior” in response: “...trying to think both of strategy and tactics, and the soldier being hit and being machine-gunned.” But the global support Ressa mustered through investigative journalism and press freedom advocacy had an international impact and that was a source of tremendous psychological benefit for Ressa, because “It showed her that there’s hope,” Gloria said.

Foreign actors

Foreign States also feature extensively in the tweets identified as containing highly explicit abuse, as well as a conspiracy theory that Ressa was in league with some of them who were allegedly paying her to spread lies. This was also a theme evident in the Facebook data. On that platform, Ressa was targeted in influence operations originating in China designed to foment popular support for political campaigns in the Philippines. That disinformation network was removed by Facebook based on evidence of 'coordinated inauthentic behavior'.

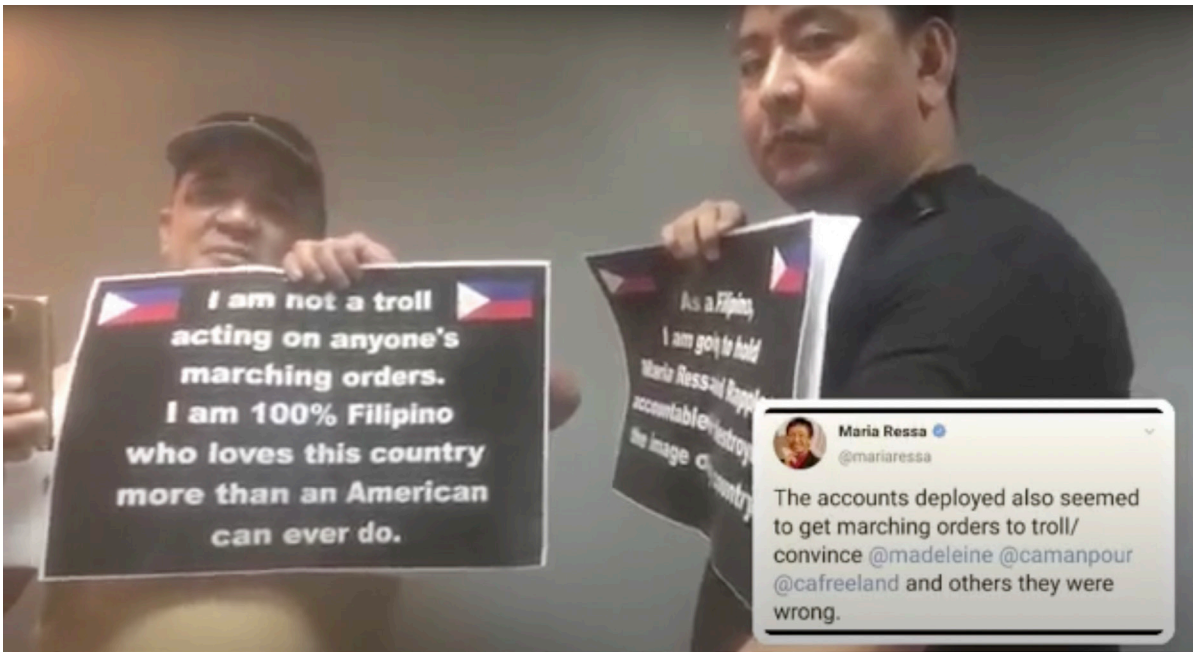
Networked gaslighting

Another feature of credibility-based attacks is the orchestration of disinformation laden 'pile-ons' aimed at Ressa, which are designed to discount her investigations into disinformation campaigns associated with political interests. These practices are features of what we called networked gaslighting. In Ressa's case, the target of the attack is falsely accused of practicing the behaviour of the attackers.

One early example involves Ressa's 2016 investigation Propaganda War: Weaponizing the Internet examining 'astroturfing'[17] and 'sock puppet networks' of fake accounts linked to President Duterte's election campaign, practices also later associated with the extrajudicial killings connected to the so-called 'drug war.' In response to Ressa's and Rappler's reporting on what they describe as "government-sponsored information operations," 'patriotic trolls' swarmed Ressa's social media zone, often prompted by partisan bloggers who encourage their followers to prove that they are not paid trolls.

"Even if some of the people who send Maria horrendous, abusive images have no known direct link to the State," Caoilfhionn Gallagher QC said, "they are likely to be doing this, or feel emboldened in doing this, because the State is sending the message that this person is fair game." These campaigns have spilled offline, increasing the physical threats Ressa is facing. In one instance, involving the doxxing of Ressa with her email and office address published online, pro-government social media activists came to the Rappler newsroom in person. They bypassed security and protested outside the glass walls of the newsroom, while holding up signs replicating some of the offensive hashtags and narratives swirling on social media.

[17] The act of manufacturing consent through influence operations designed to create the false impression of a groundswell of support within online communities.

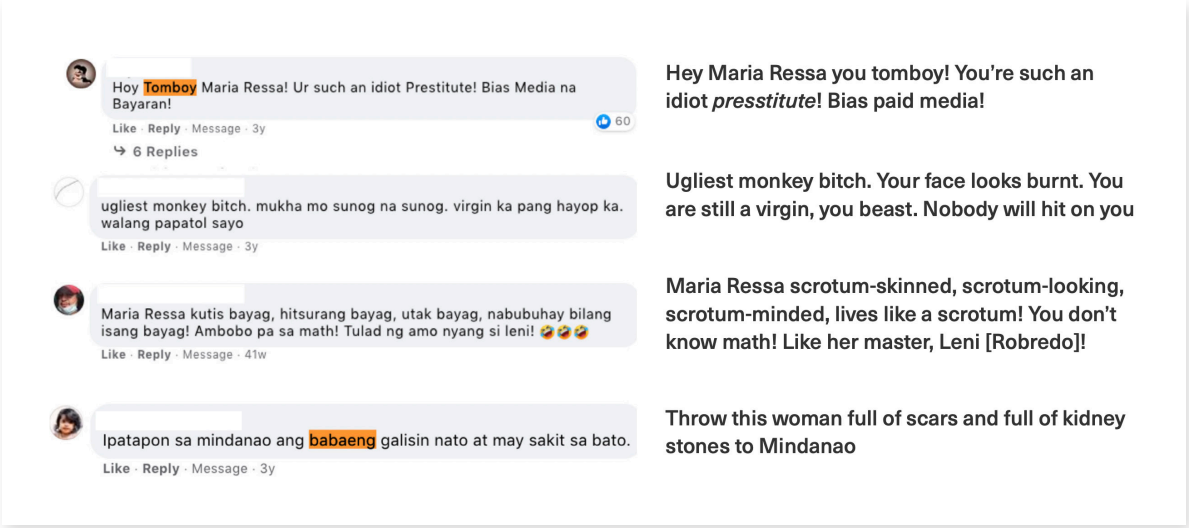


The online abuse targeting Ressa and Rappler spills offline in February 2019. These men came to the Rappler newsroom in Manila and held their 'posts' up to the glass windows of the office when they were barred entry by security (image: Rappler).

Personal attacks designed to shame, humiliate and silence

A total of 40% of all the abusive posts studied could be categorised as 'personal attacks' and these are frequently sexist, misogynistic, racist and vulgar in combination, focusing on Ressa's physical appearance. The biggest cluster of personal attacks were comments related to her face ("mukha"). She was often compared to animals like monkeys and dogs (classic racist and sexist tropes) and in several instances, her eczema was compared to a scrotum - a form of abuse which has more recently grown into a viral meme that jumps platforms. Terms like "scrotum face," "idiot," "imbecile," "bobo" (dumb or stupid), "moron," and "psycho" were prevalent.

Additionally, she was condescendingly referred to as "ang babae", or "this woman," demonstrating the sexist undertones of much of the abuse that targets her. Comments about Ressa's sexuality, including homophobic slurs (e.g., 'Tomboy' is slang for 'lesbian' in the Philippines) and outright profanities like "f**k [eds' asterisks] you" and "putang ina" (son of a bitch) were also prevalent within these clusters. More disturbingly - especially in a country which remains one of the world's deadliest for journalists - outright threats to Ressa were detected in this dataset, too, with commenters saying she should be sexually assaulted, die, or be killed.

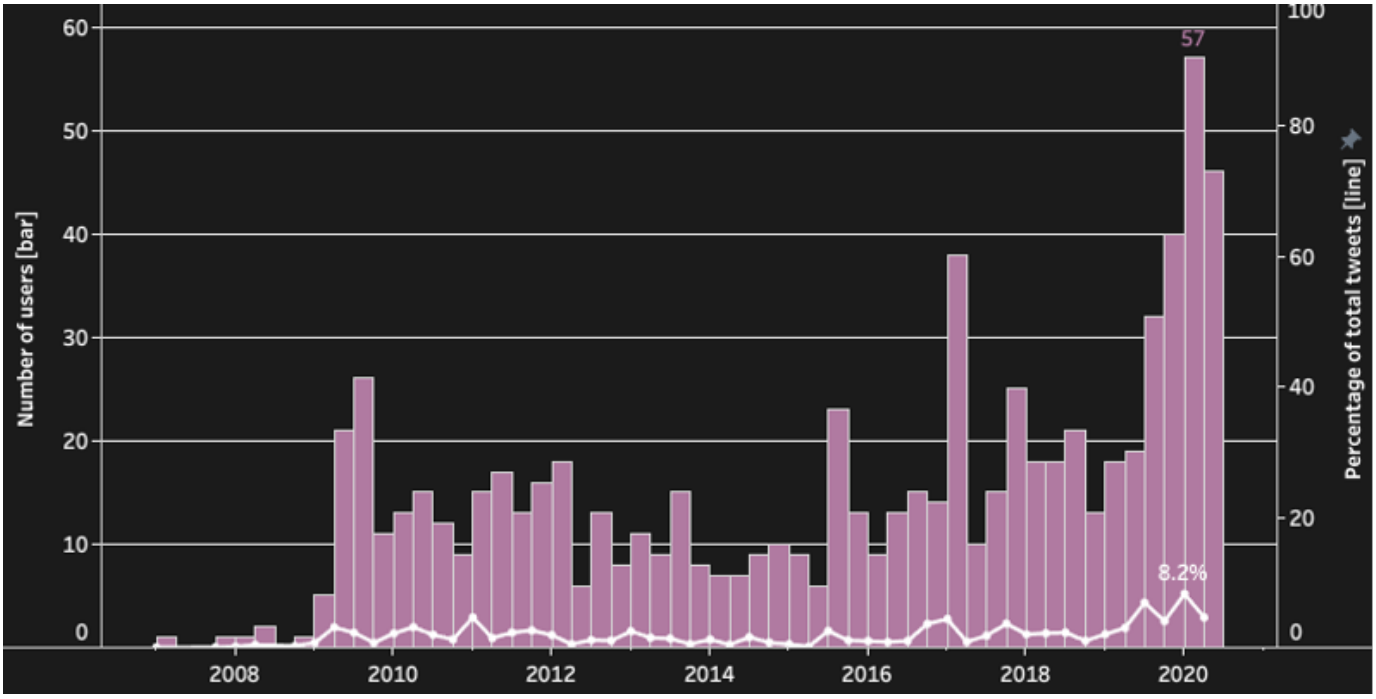


Samples of Facebook comments with personal attacks against Maria Ressa. In the comments above, Ressa is called a "lesbian" and "monkey," and her skin was compared to a scrotum. She is also often condescendingly referred as "this woman".

"I've always known online violence leads to real world violence," Ressa said. This is a view shared by Caoilfhionn Gallagher QC, who points to chilling similarities between Maria Ressa's case and that of the murdered Maltese journalist Daphne Caruana Galizia, who was brutally attacked online - with misogynistic references to 'witch burning' - before she was killed with a car bomb in 2017: "There are shocking similarities between Maria and Daphne's cases, including a long period of time in which they both experienced a combination of attacks, from multiple different sources, online and offline – State-facilitated and State-fuelled." Gallagher represents Caruana Galizia's bereaved family, who issued a statement highlighting the parallels between the cases when Ressa was convicted in 2020. "I'm aware of where this can go," Ressa said. "But at the same time, that's also why I'm very vocal. I think that the only defence is to shine the light. I continue to do my job even better. My job in Rappler is really to hold up the sky so our team can work."

iv. DEEP DIVE: NETWORK ANALYSIS OF TWITTER REACTIONS TO RESSA'S JUNE 2020 CONVICTION

According to Rappler's Executive Editor Glenda Gloria, Maria Ressa's June 2020 conviction on a criminal cyberlibel charge "...really provided the trolls [with] a powerful hashtag, because this was like a court already saying what the troll army believed and shared to be true. And so that gave them an editorial agenda - 'it's not just us saying that she's a criminal, it's the court!'"



Distribution of pro-government account creation dates for Twitter accounts targeting Ressa (source: Graphika).

In parallel with our research, data analytics company Graphika conducted a detailed analysis of 196,000 tweets from 80,886 distinct users featuring citations of @mariaressa and the term “Maria Ressa” posted between June 9th and June 17th 2020. Activity began accelerating in the days leading up to the court’s verdict, peaking on June 15th 2020, the day of the decision. While ‘mentions’ expressing shock at the verdict and support for Maria Ressa flooded the zone on the day of her conviction - including solidarity messages from international journalists and civil society organisations - dominated, pro-government accounts provided evidence of an orchestrated response. These accounts largely celebrated her conviction while attacking Ressa based on her dual Filipino-American citizenship.

Here we find strong pointers to organised “trolling”, with over 40 accounts from the pro-government segment constantly mentioning @mariaressa or citing the term “Ressa” over 30 times each within the period. Besides directly targeting her (@mariaressa), these accounts were also predominantly retweeting anti-Ressa messaging pushed by a select few accounts.

Top false narratives deployed by the pro-government 'troll army' in the immediate aftermath of Ressa's conviction were:

1 / She is now proven to be the criminal we said she was.

2 / Disinformation about the role of the State in her prosecution (i.e. they falsely argue that the case was prosecuted by a private citizen but it was a criminal prosecution waged by the State).

3 / She is a foreigner (she is a dual national), subject to foreign masters (this theme helps prosecute the false argument that Rappler is foreign-owned, which is attached to a string of cases designed to shut down the news publisher).

Pro-government accounts showed the strongest tendency to mention other users in regard to this narrative. Thus, almost 60% of these accounts referenced at least four other users within the dataset in this time period. This level of interactivity is uncommon and indicative of a possibly aligned/coordinated harassment campaign in which users aim to amplify attacks. A large number of the pro-government accounts were also fairly recently created. Accounts created during or after April 2020 were also the most active within this set - producing about 27% of the total activity (662 tweets). Finally, a significant number of these accounts have low follower counts with 5% having zero followers and a little over 25% having 10 or less followers. This combination of high activity, low follower count, and recent creation date are together possible indicators of accounts created in bulk to amplify pro-government messaging and target government critics.

v. THE ENABLING ROLE OF THE PLATFORMS IN MARIA RESSA'S ABUSE

As our big data analysis demonstrates, while a coordinated and vitriolic mob of digital aggressors instigate and fuel the attacks on Maria Ressa, Facebook - which is equated with the internet in the Philippines - is the major vector for the disinformation-laced online violence she experiences. Rappler was 'born' on Facebook and President Duterte rode the platform to victory in 2016. Ressa and the Rappler team have consistently flagged the online attacks with Facebook, which has largely done little to deal with the problem, they said. Ressa's Facebook page was overwhelmed by attacks for years, and posts that incite violence, despite violating Facebook's community standards, remained visible on the page at the time of writing. Rappler responded to the attacks with stricter comment moderation policies, but the abuse on Ressa's page only dropped when she wound down posting in the second half of 2018.

In September 2018, Ressa told Facebook executives: "If you don't change what you're doing, I could go to jail". After she was convicted in June 2020, she apportioned some blame to the company for her conviction, and she continues to argue that Facebook has enabled the destruction of democracy in the country. Ressa is very critical of Facebook's moderation policies and practices, saying that their automated reporting systems just do not work when it comes to dealing with online violence against women journalists. "I have very rarely had anything taken down when trying to report attacks using Facebook's standard online reporting system," she said.

"The design of social media turned 'wisdom of the crowds' into the mob. It's the chaos of a mob. And beyond that, it's actually pumping hate into the system," Ressa said. She believes the platforms need radical renovation - of business models and design - to stop the toxicity that overruns them. "I don't think anything is possible until we clean up the information ecosystem, until you stop the virus of lies," she said. "It's a perfect comparison to the COVID-19 virus, because it is very contagious. And once you're infected, you become impervious to facts."

Twitter is also a significant distributor of abuse against Ressa, although she says she feels "much safer" on that platform - especially since the company appears, in her view, to have begun working harder over the past few years to protect women journalists and human rights defenders among its users. "Their reporting tool aggregates similar tweets, takes less time, and is far more effective in takedowns," Ressa said.

She has long spoken about being the 'canary in the coalmine,' warning that the Philippines' information ecosystem represents the West's dystopian future. She is extremely pessimistic about the prospect of the platforms responding quickly and effectively enough without changes to accountability and liability:

> ***"The only times Facebook has done something about the attacks against me is when I have gone directly to people I know inside the company, many of whom have nothing to do with content moderation! Their systems need to be clearer, faster and more responsive to the impacts of their inaction..."***

Ressa continues to work at the intersection of investigative journalism, research and policy to respond proactively to the information ecosystem crisis. And Rappler continues to be a Facebook fact-checking partner in the Philippines, and a collaborator on investigations into disinformation networks which, in some cases, have resulted in the company removing clusters of inauthentic accounts from the platform.

The only way it will stop is when the platforms are held to account...because they allow it. It's kind of like if you slip on the icy sidewalk of a house in the US, you can sue the owner of the house. Well, this is the same thing. They have enabled these attacks. They've certainly changed my life in many ways.

Maria Ressa



The networked gaslighting of a high-impact investigative reporter

Carole Cadwalladr

I'm not biased. I'm furious. I'm boiling with rage. The bullies are winning. Lies are winning. This assault on truth, justice, democracy is winning. And we can't even see it. That video[18] – created by a British political organisation, facilitated by a global technology platform – will have an impact on other women. On other journalists. It's another line crossed.

Carole Cadwalladr, 2017

[18] See later discussion - Cadwalladr is referring to a deep fake video in which her head is superimposed on the body of a woman being repeatedly slapped by a line-up of men, one of whom holds a gun.

Carole Cadwalladr is a multi-award winning British journalist whose investigative work exposed the Facebook-Cambridge Analytica scandal that ultimately led to the biggest fines in history being imposed on the social media giant by privacy regulators on both sides of the Atlantic. It also resulted in the collapse of the "sketchy" data analytics company closely associated with former US President Donald Trump's successful 2016 election campaign. But from the moment *The Guardian* and *Observer*[19] newspapers published the first story in Cadwalladr's ongoing investigative series into the scandal that compromised the Facebook accounts of up to 87 million users, the journalist became the target of a malign, misogynistic, disinformation-laced campaign of online violence which has grown increasingly threatening over time. This campaign has also created the enabling environment for her ongoing legal harassment by political actors.

Cadwalladr's reportage linked the Cambridge Analytica scandal to both the election of former US President Donald Trump, and Brexit - the referendum which led to the withdrawal of the UK from the European Union. It suggested the widespread manipulation of Facebook users by political actors using microtargeting techniques, which fed into highly divisive politics in the US and the UK between 2016 and 2020. And continues to resonate. Her journalism has led to criminal investigations and parliamentary inquiries in multiple countries.

Despite reporting extensively on the global impacts of Facebook privacy breaches and disinformation on the platform, Cadwalladr barely uses Facebook, she said. Twitter has been the main vector for the cascading gendered abuse, threats and harassment that she endures. But the fuel is provided by pro-Brexit political actors and donors, whose abuse has been amplified by right-wing media, and even a prominent (now former) BBC Politics presenter. The objective, she said, is to hold her up to ridicule, discredit her, and thereby erode trust in her accountability journalism:

"One of the most important things is the way that credentialled people become part of it. For me, it's right-wing journalists and 'trolling' MPs. That's when it becomes really hard to counteract. From the professional point of view, you're perceived like a sort of divisive, controversial figure. Something which for your male colleague would have passed without comment. But it's making you into a controversial figure, and I think this is very effective."

[19] The *Observer* is *The Guardian*'s companion Sunday national newspaper. Both papers are published by Guardian News Media, along with the Guardian Weekly.

In this interdisciplinary case study, we conducted a big data analysis of nearly 2.1 million English language tweets directed at Carole Cadwalladr from December 1st, 2019 to January 14th, 2021. Applying the same Natural Language Processing (NLP) techniques we used to analyse the data in the Maria Ressa case study[20] featured in this report, we isolated 10,400 tweets identified by Artificial Intelligence (AI) tools as clearly abusive and subjected them to granular analysis. We undertook this quantitative research in parallel with longform qualitative interviews conducted with Cadwalladr, *Observer* editor Paul Webster, and Guardian Media Group's Director of Editorial Legal Services Gill Phillips to understand the impacts of the online violence Cadwalladr experiences, and the responses to it. We then created a timeline of events that helped us map her experience of sustained online violence over the past four years.

According to our research, the main goal of the abuse levelled at Cadwalladr is to discredit her professionally, thereby undermining trust in her critical reporting of the Cambridge Analytica scandal and its aftershocks, including questions regarding the accountability of the platforms for disinformation and hate speech. The parallel purpose is to discredit the journalism of *The Guardian* and *Observer* newspapers. "If you report on disinformation, you become a target of disinformation," she said. "If you report on the far-right, you become a target of the far-right."

"I am a cultural war hate figure"

The nature of the abuse levelled at Cadwalladr is highly gendered, and at times misogynistic. Most typically, she is maligned and dismissed as a "mad cat lady," mocked for being middle-aged and childless; labelled "crazy", "hysterical" and a "conspiracy theorist". All this is "really humiliating", she said.

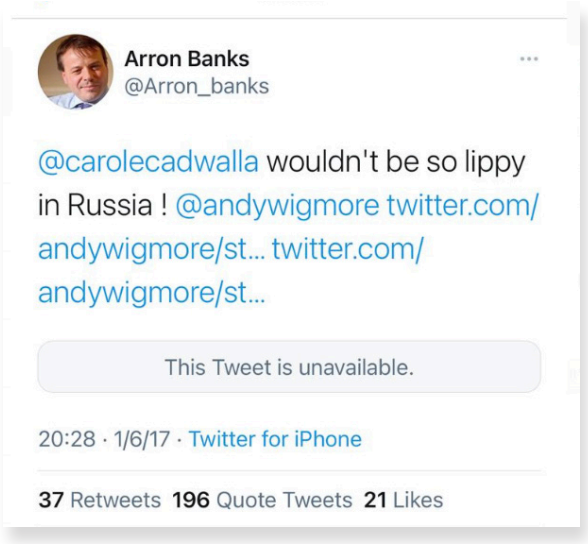
The *Observer*'s editor Paul Webster has witnessed the impacts on Cadwalladr: "Carole has been subjected to a fairly sustained and aggressive fusillade of online attacks in her reporting. She's a controversial writer who's chosen a series of very contested areas to report on, and she has been subjected to a great deal of online abuse." Cadwalladr's free-wheeling social media style and her refusal to be silent make her a bigger target. And when she posts controversial comments or occasionally corrects errors on Twitter, the abuse becomes torrential. To Cadwalladr, this feels like the digital equivalent of a mob attack on an "unrespectable" woman. "A few hundred years ago I would have been burned at the stake," she said.

The abuse has not been contained to the online realm. Cadwalladr was physically stalked in 2018 by a "shady" man with a military and cyber espionage background who first tried to befriend her, and then began sending threatening text messages. She has also been hit with multiple defamation claims by one of the subjects of her investigative reporting, a wealthy businessman who is also a prime instigator of the online harassment she experiences, Leave.EU (the unofficial pro-Brexit

[20] See the detailed methodology laid out in the full-length Maria Ressa big data case study here: https://www.icfj.org/our-work/maria-ressa-big-data-analysis

campaign) founder Arron Banks. Some of the tweets he directs at Cadwalladr have been received as menacing. They are certainly highly gendered. The defamation cases filed by Banks against Cadwalladr have been declared Strategic Lawsuits Against Public Participation (SLAPPs) by a coalition of eight freedom of expression groups led by Reporters Without Borders (RSF). They described the defamation actions as: "vexatious in nature and intended to silence Cadwalladr's courageous investigative journalism."

The *Guardian* and *Observer* have also been threatened with legal action in the course of publishing Cadwalladr's investigative reporting into Cambridge Analytica - including by Google and Facebook. However, Banks' defamation claims target her as an individual - a typical characteristic of SLAPP suits. They respond to statements she made during a 2019 TED Talk and a tweet in which she shared the talk.



This tweet from Arron Banks quotes a tweet from the communications director (Andy Wigmore) of the unofficial Brexit campaign Leave.EU. The Wigmore Twitter account is currently suspended.

These factors, combined with her status as a freelance columnist at the *Observer* rather than a staff reporter, have left her personally exposed and responsible for her own legal defence. "The thing which I think is important...is that the lawsuit that I have is within this context of a four-year campaign of harassment, intimidation, threats of violence... It was kind of like falling down the 'rabbit hole'." In her view, she was a threat to these very powerful actors.

While the Guardian Media Group (which owns the *Observer*) is not funding Cadwalladr's defence in the defamation cases, her editors have publicly defended her reporting through statements like this:

40%

of obvious abuse in the 'personal attacks' category was sexist and misogynistic

> ***“Carole’s brave reporting has made waves around the world, and given the public much more insight into the secretive ways some powerful people and organisations have sought to influence our democracies… This case is a very worrying example of a wealthy person singling out an individual journalist, and using the law to stifle legitimate debate and silence public interest journalism.”***

In the past year, Cadwalladr’s experience of online abuse and harassment has also been influenced by the coronavirus pandemic. “My Brexit trolls converted overnight into Corona trolls,” she said. “I am a cultural war hate figure and so at the start of the pandemic, when I spoke up about the government’s decision to not lock down based on scientific evidence it was refusing to publish, it became very acute.”

Her fightback strategy in the face of sustained online violence has involved countering disinformation through investigative journalism, forming support networks with other women journalists experiencing social media abuse, and launching civil society collaborations designed to hold social media companies and political actors to account. It is a “solutions focus”, Cadwalladr said. These collaborations include a network of scientists critical of the UK’s response to COVID-19 and the ‘Real Facebook Oversight Board’ - both initiatives of grassroots advocacy group, The Citizens. She stated: “I’m just trying to process what I’ve gone through for the last four years… I’m trying to look structurally at the news and information ecosystem and work out how we can counteract disinformation, how we can help others who are experiencing this. I really have had enough of being the national punching bag.”

Cadwalladr recently recruited Maria Ressa - the subject of our other big data case study - to sit on the ‘Real Facebook Oversight Board’. And while their cases are not directly comparable, there are some noteworthy similarities in the patterns of attack, and the role of gendered online violence against a journalist in creating the enabling environment for legal harassment. In both cases, ‘patriotic trolling’, niche bloggers/ influencers and partisan media figures fuel the attacks, but domestic political actors, not foreign States, are the biggest source of the online violence they experience. As Cadwalladr points out: “There is this misconception about what online attacks are - this idea of these kind of faceless foreign bots. Whereas for me, the particularly pernicious and troublesome thing is this right-wing information system that is all powerful in Britain, working at all of these different levels.”

12 Key Findings

Types of abuse levelled against Cadwalladr:

55% of obvious abuse detected targeted Cadwalladr occurs at the personal level. It was highly gendered and designed to hold her up to ridicule, humiliate, belittle and discredit.

40% of the abuse was categorised as harassment designed to undermine Cadwalladr's professional credibility and trust in her journalism.

21% of all obvious abuse levelled at her was sexist, misogynistic or sexually explicit.

5% of the abuse was politically-based.

Characteristics of abuse against Cadwalladr:

- The online violence Carole Cadwalladr experiences is a feature of the enabling environment for her offline legal harassment.
- The abuse was constant and sustained, with several peaks per month delivering intense abuse.
- The cumulative impacts of the sustained online abuse, harassment and attacks over a four year period have created a gaslighting effect, chilling Cadwalladr's investigations and delivering deep personal impacts.

Abuse tactics and triggers:

- **Disinformation tactics are clearly in evidence**

- **There is obvious evidence of trolling behaviours among abusers, with key identifiable political instigators leading the pack.**

- **The online violence experienced by Cadwalladr has been instigated by subjects of her investigative reporting, amplified by fringe right-wing media, worsened by conservative columnists, and inflamed by the sexist behaviour of high profile establishment journalists.**

- **Cadwalladr's internationally consequential, multi-award winning journalism focused on powerful individuals and entities are what made her a prime target for online violence.**

- **Cadwalladr is disproportionately attacked for making corrections and contentious comments, as the spikes of abuse in our Twitter data demonstrate.**

The gaslighting effect

Cadwalladr has been subjected to four years of deeply sexist and misogynistic online violence that takes the form of a constant wave of abuse, with several crests each month. This pattern of abuse - building gradually over time - appears designed to destroy the confidence of the target and undermine their credibility without appearing overtly offensive on a tweet-by-tweet basis. This method typically operates via an organised, or semi-organised, network of abusers for greatest effect (the 'pile-on', 'dogpiling', or 'brigading' approach), and can lead to echo chambers of abuse, where the same abusive message is retweeted many times. We see these methods clearly in play in the case of Cadwalladr. Her abusers are frequently interlinked - organisationally, or through association with pro-Brexit rhetoric. We call this networked gaslighting.

40%

of abuse is aimed at undermining Cadwalladr's journalistic credibility, as well as her integrity.

i. ANALYSIS: HYSTERICAL HAG, STUPID BITCH, CRAZY CAT LADY, SHUT UP!

We identified 10,400 separate instances of obvious abuse against Cadwalladr in our Twitter dataset which consists of 2.1 million English language tweets collected between December 2019 and January 2021.

Sexist, misogynistic, and explicit abuse represented 40% of obvious abuse in the 'personal attacks' category and 21% of abuse detected overall. It is notable, too, that the pile-on of abuse Cadwalladr experienced included significant elements of anti-journalism and anti-mainstream news media rhetoric, reflecting the demonisation of the press on the global stage, and the weaponisation of terms like "fake news" to chill critical reporting.

The dominant abusive phrases and terms illustrate the gendered nature of the online violence Cadwalladr experienced during the period studied (e.g, variations on "stupid woman", "crazy cat lady", and "witch") which is clearly designed to belittle and humiliate her, while also eroding trust in her reporting. The main objective appears to be to intimidate her into silence.

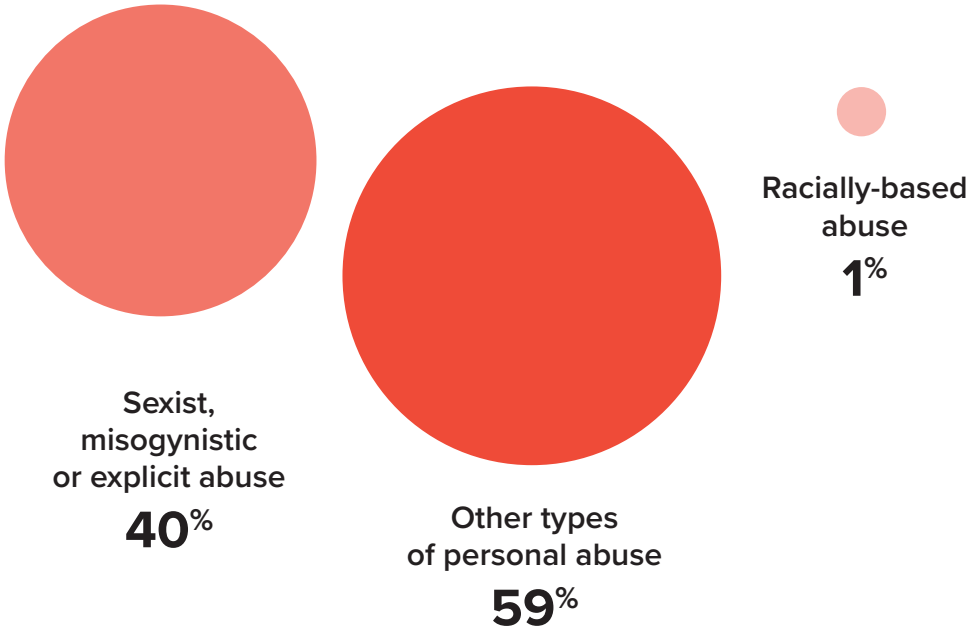
25%

of the credibility-related abuse involves terms implying that Cadwalladr is stupid or mentally ill in some form.



The word cloud above shows the most frequently occurring abusive terms tweeted at Carole Cadwalladr (occurring at least 20 times), normalised by case.

Next, after analysing the clearly identifiable abuse in the 'personal' category (55% of all abuse), we classified it into three subcategories:



The other 40% of the abuse that we identified is aimed at undermining Cadwalladr's journalistic credibility, especially her alleged "stupidity" as well as her integrity. Typical slurs of this kind include labelling her a "liar" or claiming she is talking "crap", "bullshit", or "bollocks". The objective of this abuse is to scare her away. She is frequently told to "f**k [eds' asterisks] off". She is often referred to as "Carole Codswallop" or "Codswallop Cadwalladr", again insinuating that she 'talks rubbish' and her journalism is not to be trusted. But it is the sustained, low-intensity, high-frequency, high-volume nature of the attacks Carole Cadwalladr experiences, not only the content of the abuse, that can be understood to be so cumulatively damaging.

Looking at breakdown by frequency, we established that over 25% of the credibility-related abuse involves terms implying that Cadwalladr is stupid or mentally ill in some form (e.g., "idiot", "twat", "moron", "cretin", "silly", "crazy", "fool", "nutter"), with the term "liar" also appearing over 300 times, alongside terms like "shut up", "STFU" and "f**k [eds' asterisks] off". This language is specifically designed to undermine Cadwalladr's journalistic reputation and professional credibility while also seeking to silence her. 'Sexist, misogynistic, and explicit' terms include examples such as "dickhead", "f**k [eds] off", "cunt", "witch", and "hag", along with terms involving sexual acts and intimate body parts.

Instances of political abuse frequently involve anti-Brexit sentiment such as calling her a "remoaner" (a pejorative term for people who supported the campaign for the UK to remain in the EU). Other insults included political abuse such as "Common Purpose globalist whores", "F**k [eds' asterisks] off you commie twat" and frequent use of "Libtard" (a pejorative term for a liberal thinker) which categorised Cadwalladr as left of centre and lacking objectivity - another way to discredit a journalist in the UK context. Some of the worst abuse levelled against Carole Cadwalladr in this category during the period of this study involved people wishing she would die, as illustrated by this tweet:

@carolecadwalla The only failure is scum like you. You [sic] still throwing your toys out the pram that we left the EU. Hope you get covid and die from it [21]

Many tweets using such hashtags combined misogynistic language with pejorative descriptions of the mainstream press. Here is an example of a tweet sent to Cadwalladr after she tweeted a link to a story about this UNESCO-commissioned research into online violence against women journalists, ahead of her appearance on a 2020 World Press Freedom Conference panel:

[21] This tweet, from a now suspended Twitter account, was published November 20th 2020.

@carolecadwalla @julieposetti Female? Male? All #scummedia should be treated accordingly - there should be no discrimination when it comes to destroying all you Common Purpose globalist whores. Is there anything lower than you and your ilk? #presstitute #evil

This now deleted tweet, which also tagged this study's lead author (i.e. it radiated to capture affiliated targets) features the familiar hashtag associated internationally with gendered online violence against women journalists - #presstitute.

According to our analysis, the main themes associated with abuse against Carole Cadwalladr were: democracy; public health and COVID-19; Brexit, Europe and immigration; and foreign affairs. These are also the issues that Cadwalladr most frequently wrote about and commented on during the period in focus. But one of the most striking features of the abuse against her is the deployment of misogynistic tropes. She is frequently labelled "mad" and "hysterical" by critics who emphasise her age, relationship status and childlessness, and question her desirability. This sexist 'othering' is a way to treat her as a deviation from the norm. It has also been amplified by some very high-profile figures in the UK press, in addition to popular right-wing blogs. To Cadwalladr, this has all the hallmarks of a witch hunt.

"Codswallop", the "crazy cat lady" and the former BBC presenter

We found 2,921 mentions of the term "cat lady" and its various manifestations (e.g. "cat woman") in our dataset of obvious abuse against Cadwalladr. It is one of the most pernicious and pervasive forms of abuse in evidence, and it is usually used in conjunction with "mad" or "crazy." "I would still say that to this day Andrew Neil is largely responsible for the ongoing misogynistic abuse and threats that I get," Cadwalladr said. She is referring to a now deleted tweet from the then BBC Politics presenter and chair of the parent company of the conservative magazine *The Spectator*, Andrew Neil. In November 2018, Neil tweeted about Cadwalladr, branding her a "mad cat woman" and ridiculing her as "Karol Kodswallop", insinuating that she resembled a character from the satirical cartoon The Simpsons.



Business Insider's Senior Politics Editor Adam Bienkov[22] defends Cadwalladr after she receives sexist abuse on Twitter from Andrew Neil.



Cadwalladr tweets about the abuse she received following Andrew Neil's 2018 tweet[23].

[22] https://twitter.com/AdamBienkov/status/1062295554946932736?s=20
[23] https://twitter.com/carolecadwalla/status/1064074315828744192?s=20

According to our research, the Leave.EU donor and founder Arron Banks - who, as described above, is currently pursuing SLAPP-style defamation claims against Cadwalladr - was the first Twitter user to disparage the journalist as a “crazy cat lady”. He also frequently labelled her “Carole Codswallop”. This language was used to denigrate her as she pursued stories probing possible links between the Cambridge Analytica scandal and Leave.EU, the unofficial pro-Brexit campaign that Banks funded. Ten days before Neil called Cadwalladr a “mad cat woman”, she wrote in The Guardian that since her investigative reporting began to focus on Leave.EU and Banks, he had called her: “...hysterical, insane, a lunatic, a mad woman, a conspirator, a loony, a mad cat lady, a nasty piece of work, a criminal, a bully, a mad cat lady, a loony, a tinfoil hat nutter, a hacker, a mad cat lady, a loony, a bitter Remoaner, a lone conspiracy theorist, an enemy of the people.”



Cadwalladr says she is a target of a toxic right-wing media ecosystem in the UK.

This abuse mirrors a pattern of online violence against women journalists that is now recognisable internationally. This involves a process of instigating, amplifying and legitimising gendered online harassment, abuse and attacks. In Cadwalladr's case, the sexist and misogynistic tropes deployed by Banks and his supporters were amplified by the popular right-wing blog Guido Fawkes, and then legitimised by sympathetic voices in the mainstream news media. Within this right-wing information ecosystem, we see abusive content about Cadwalladr being cross-pollinated through retweeting and quote-tweeting by various actors affiliated with, or sympathetic to, the campaign for Britain's exit from the EU, and/or disparaging towards Cadwalladr's reporting. Notable among them during 2017/2018 are Arron Banks, Leave.EU's Twitter account, Leave.EU's Communications Director Andy Wigmore (who had been suspended from Twitter at the time of writing), the Guido Fawkes blog, a foreign State actor, *The Spectator*, and its holding company chair, Andrew Neil.

Neil deleted his "mad cat woman" tweet following a backlash but he did not apologise. The BBC's corporate communications team did, however, issue a tweet in which they stated that Neil recognised the tweet was "inappropriate". The following day, the then BBC Director General Tony Hall told a conference:

"On Twitter there are constant anonymous threats to journalists simply reporting on opinions that some people might not want to hear. Some of the material that journalists have had to face is quite frankly disgraceful. It is an attempt to intimidate people and stop them doing their jobs. For the sake of all journalists - we need to defend our role - seeking out the facts, no matter how inconvenient they may be for others. Because journalism matters - whether you're in broadcasting, in the press or working online."

The next week, *The Guardian* reported that a number of senior women journalists at the BBC had complained to executives about Neil's tweet, suggesting inequitable treatment regarding social media policy enforcement. A BBC spokesperson was quoted as saying: "The tweet was sent from Andrew's personal account, however the BBC has social media guidelines which it expects all staff to follow and these have been discussed with Andrew." Meanwhile, the misogynistic insult continues to reverberate: "I am now called a 'mad cat lady' dozens of times a day on social media, every day," Cadwalladr said. In addition, "Codswallop", and Neil's variant "Kodswallop", are also pejorative nicknames still in regular use in abusive tweets against Cadwalladr according to our research.

ii. WHO IS SENDING THE ABUSIVE TWEETS TARGETING CAROLE CADWALLADR?

Of the 10,400 obviously abusive tweets identified in our dataset, there are 7,744 unique authors. Of the serial abuse senders, the most prolific sent 38 obviously abusive tweets, while 155 tweeters sent five or more that our NLP tool categorised as abusive. Only two of the top eight most prolific abuse senders in our dataset still have live Twitter accounts. Of the other six, three have been deleted and three were suspended at the time of writing. Leave.EU's Director of Communications Andy Wigmore was the Twitter user who abused Cadwalladr most prolifically during the period. He sent 38 highly abusive tweets before his account was suspended on November 8th, 2020. Here is one of the abusive tweets Wigmore sent before his suspension, in which we see him allude to SLAPP suits and tag both Andrew Neil (who triggered the "mad cat woman" pile-on of 2018) and Arron Banks, who is suing Cadwalladr for defamation:

> ***Pointing out facts by @afneil (a proper journalist) must be like a SLAPP round the chops or scraping your nails down a chalk board Codswallop @carolecadwalla get used to it more facts coming your way. @Arron_banks***

The second most abusive account was an anonymous Twitter account which also overlapped with another anonymous account using the same profile photo and an almost identical Twitter handle. Both accounts were suspended in late 2020. Out of the 7,744 authors of obviously abusive tweets in the dataset, as of March 15 2021, 943 of these had deleted their accounts, and 1,021 had their accounts suspended. In total, this means that just over 25% of the accounts are no longer active. This suggests that those accounts abusing Cadwalladr most prolifically are people whose misconduct on the platform has been relatively extreme, generating a high removal response rate from Twitter. Authors of abusive tweets also have more recently established accounts than tweet authors in the non-abusive set, fewer followers, follow fewer users, and they post slightly fewer tweets. Viewed together, this data pertaining to the behaviour of the Twitter users in our dataset who were obviously abusive towards Cadwalladr during the period under examination indicates clear patterns of trolling behaviour with identifiable instigators and 'ring leaders'.

iii. ABUSE SPIKES AND TRIGGERS

Cadwalladr is an outspoken and reactive Twitter user, reflecting her status as a high-profile freelance columnist at the *Observer*, and meaning that she is relatively unfettered in her social media conduct compared to journalists bound by corporate social media policies. At the same time, she is also a woman and a journalist who feels battered after years of exposure to unrelenting online violence. “It’s changed my life,” Cadwalladr said. Early on, she was hacked; she has been stalked; she perceives that the trolling has ruined her reputation within the UK news industry; she acknowledges her reporting has been chilled; she has experienced the trend of victim-blaming - judged for “having answered back on Twitter” - which is slowly muting her social media use; and two of the whistleblowers at the heart of her Cambridge Analytica investigations were also targeted in coordinated online and news media attacks designed to undermine their credibility, she said.

All of this was going on as Cadwalladr was winning a cache of premier British journalism awards, such as the 2018 Orwell political journalism prize, and recognised as a Pulitzer Prize finalist with the *New York Times*. And now, she faces possible bankruptcy, as defamation actions brought by one of the powerful subjects of her award-winning investigative reporting wend their way through the courts. “I’m processing actually what it’s like being under attack for four years. At various points, I think I have overreacted to things as well, you know. You’re so used to being attacked that you respond like you are on a bit of a trigger.”

Her realtime reflections are weaponised against her, as are any realtime errors. Cadwalladr corrects the relatively few mistakes she does make but she is offered no forgiveness, nor shown any grace on Twitter by her dedicated detractors and trolls. Our dataset demonstrates a continuous ebb and flow pattern of abuse, frequently associated with a backlash against her commentary, with two to three attack spikes per month. This sort of digital gaslighting - a long-range persistent pattern of abuse - is designed to wear the target down.

As a woman covering the global political, social and human rights ramifications of big technology and viral disinformation, and historically divisive political movements (on both sides of the Atlantic); who dared to break major investigative stories as a features writer and columnist, Cadwalladr could be seen as a natural lightning rod for online abuse in the toxic communications ecosystem of the early 21st century. But while her reporting and tweets are often cast as “controversial”, the abuse she receives is an entirely disproportionate response to her occasional errors and her engaging, reflexive, opinionated and consequential writing.

iv. THE ROLE OF *THE GUARDIAN* AND *OBSERVER*

The online violence Carole Cadwalladr is subjected to is designed to silence her, chill her critical journalism and discredit those who publish it, namely: *The Guardian* and *Observer*. In August 2020, Guardian Media Group took legal action to shut down a website generating fake Guardian headlines and byline profiles which were being shared with the trending Twitter hashtag #TrollingTheGuardian. Carole Cadwalladr was one of the journalists targeted by the now defunct guardianmeme.com website. But the action was triggered by a tweet from a UK member of parliament who was similarly attacked after she wrote a *Guardian* column. She tweeted that the site put her in danger, along with her family and staff.

Guardian Media Group's decision not to fund Cadwalladr's defence in the ongoing defamation cases - the enabling environment for which includes the online violence she experiences - is a source of grief and frustration for Cadwalladr, who regularly points out that she could "lose [her] house" as a result of the litigation. The company's position is that because the defamation claims are unfortunately made against Cadwalladr individually, in relation to speeches and a tweet, not reports published by the outlets, they are not a party to the legal action, and they are unwilling to set a precedent by funding Cadwalladr's defence. "If she had been sued individually over something she wrote for us, we would be all over it. But if we do this for her, where do we stop?" Editorial Legal Director Gill Phillips asked. "So we've had to say, really regrettably, we cannot financially support this, but we will give you whatever support we can morally, ethically, including every time we do anything on SLAPPs. And we will continue to defend the journalism she does for us."

Cadwalladr's response has been to crowdfund her defence. To date, she has raised several hundred thousand pounds from supporters. She is also working proactively to advance protocols for dealing with online violence against women journalists at *The Guardian* and *Observer* - emphasising the function of disinformation campaigns in such attacks, along with the experiences of more isolated freelancers. For Cadwalladr, dealing with online violence needs to be at least a two-way street within news organisations:

> ***"As much as they want rules about how journalists should behave online, I also think that there should be rules about how the organisation should behave towards its journalists. They need some sort of rapid response unit. They need to have actions that they put into play when one of their journalists is attacked - that there are obligations which come to them, too."***

v. THE ROLE OF THE PLATFORMS

Cadwalladr says she is very frustrated with the big tech companies that act as vectors for online violence, viral disinformation and privacy erosion. Facebook and Google have been the focus of her investigative journalism, while the online attacks she experiences are facilitated by Twitter. “I report stuff all the time on Twitter and they never ever do anything. They never take it down,” she said. “There’s these endless spoof accounts...I’m always reporting stuff and I don’t think anybody’s ever successfully had one upheld.” She was particularly aggrieved by Twitter’s failure to remove a Leave.EU deep fake video in which she was depicted being repeatedly slapped. The video, which featured one man with a gun and one with a hammer, was clearly “incitement to violence,” in Cadwalladr’s view. In the end, it was *The Guardian*’s intervention in contacting the source of the post that resulted in its removal, 48 hours later, she said.

Believing that reporting, blocking, deleting and muting are acts of futility against self-regenerating troll armies, Cadwalladr has turned to grassroots campaigning and the development of loose support networks through her work at The Citizens. Interestingly, this has involved working with data scientists she has met on Twitter to map networks of abuse and develop response mechanisms to support other users under attack.

Concerning Facebook and the impunity with which the company has been able to act, Cadwalladr said representatives repeatedly lied to her in the course of the Cambridge Analytica investigation, and the company is just too big to hold accountable:

> ***“ The FTC fined them a record US$5 billion but it had no impact whatsoever. Their share price actually went up...because there is no mechanism to hold them to account. The only accountability structure we have has no effect because the platforms are too big. You have got to find other ways, and more creative ways essentially. ”***

Her alternative approach has been to launch the 'Real Facebook Oversight Board' - a collection of academics, civil society experts, social media critics and prominent journalists who do not have faith that the Facebook-funded official Oversight Board and its limited remit will do the kind of urgent accountability work needed, with appropriate transparency, and at the scale required.

One thing that needs to change within the social media companies is resourcing to deal with online abuse and harassment, according to some. "They need far more people," *Observer* Editor Paul Webster said. "They need to recognise their roles and responsibility as publishers and act accordingly. So, curate that material in the way that we do - take legal responsibility for it. So, if somebody is libelled in tweets and in online posts, then you have recourse to the people who carry the material as well as the people who make the libels."

While *The Guardian* has had some limited success in getting the platforms to remove abusive content in certain cases, and deplatform individual offenders in others, the scale of the problem is not able to be managed with present systems, according to Guardian Media Group's Director of Editorial Legal Services Gill Phillips: "Look, of course, we can write letters and ask people to stop it, and point out it's horrible. [But] the only thing that can stop it is the people who are giving them the platforms that allow them to make these attacks."



Conclusion and recommendations

This report demonstrates that online violence against women journalists is a global phenomenon, albeit one with uneven impacts that are heightened at various intersectional points, including racism, religious bigotry, sectarianism, homophobia and disinformation. There is a climate of impunity surrounding online attacks on women journalists which must be more urgently and effectively addressed because impunity emboldens the perpetrators, demoralises the victim, erodes the foundations of journalism, and undermines freedom of expression. For too long, the emphasis has been on making women journalists responsible for their own defence and protection, rather than making the perpetrators and instigators, the platform enablers, and law enforcement and media employers accountable.

What more can be done?

These methods of attack are growing more sophisticated, and they are evolving with technology. They are also increasingly networked and fuelled by political actors. This points to the need for responses to online violence to grow equally in technological sophistication and collaborative coordination. Another point highlighted by our research: most women journalists do not report or make public the online attacks they experience, in line with low levels of reporting when it comes to violence against women more broadly. As our research participants also demonstrated, many media employers still appear reluctant to take online violence seriously. This aligns with the evident failure of the internet communications companies - whose social networks, messaging and search services facilitate much of the harassment, intimidation, abuse and threats targeting women journalists - to take effective action to address this freedom of expression and gender equality crisis.

28 recommendations for action

Cross-cutting recommendations for all stakeholders:

- Recognise that the problem of online violence is transnational, and operates in the context of huge and profitable technology companies.

- Facilitate and encourage coordinated, global multi-stakeholder cooperation and exchange of good practice between States, and across professional boundaries, in the interests of effective implementation of holistic measures for tackling online violence against women journalists in all countries.

- Foster and fund collaborative responses involving civil society organisations, journalists' networks, and researchers to gain more granular knowledge about targeted online violence, and to develop collective responses to protect and support women journalists.

- Invest in genuinely independent research into the fast-moving nature and scale of online violence and social media company responses, addressing new platforms (including those providing entertainment and gaming services) and coordinated cross-platform trolling.

- Recognise the intersectional threats associated with gendered online violence, such as racism, religious bigotry and homophobia, and respond accordingly - through policy development and training.

Intergovernmental organisations, including UNESCO, could:

- Ensure that the mechanisms and protocols developed to safeguard journalists and end impunity are appropriate for cases of orchestrated online violence against women journalists.

- Recognise and work to counter the role of officials active in facilitating and orchestrating large-scale and continuous online attacks on women journalists.

- Monitor, research and record evidence, and speak out against, online violence associated with crimes against journalists in a gender-responsive and gender-disaggregated manner.

- Consider initiating a multi-stakeholder 'early warning system' to trigger interventions (including from UN Special Rapporteurs) in cases where there is a significant risk to a target under attack online.

- Develop and provide training and education for law makers, law enforcement agencies and the judiciary to enable them to deal more effectively and appropriately with online violence against women journalists.

Individual States could:

- Ensure that laws and rights designed to protect women journalists offline are applied equally online as required by UN GA Resolution A/C.3/72/L.35/Rev.1 (2017) and reaffirmed by UN GA A/RES/74/157 (2019), which calls on States to observe the particularities of online threats and harassment of women journalists through: "Collecting and analysing concrete quantitative and qualitative data on online and offline attacks or violence against journalists, that are disaggregated by, among other factors, sex" and "...publicly and systematically condemning online and offline attacks, harassment and violence against journalists and media workers."

- Consider introducing protocols and guidelines to act against officials who engage in gendered online violence and ensure prosecution of those who attack women journalists.

- Make social media companies more clearly accountable for combating online violence against women journalists.

- Introduce regulation that provides victims of online violence with effective access to appeals against platform (in) action, including (where required) an independent, national ombuds facility.

Political parties and other political actors could:

- Desist from mounting attacks (on and offline) against women journalists and discourage pile-ons against them.

- Punish members and officials who partake in acts of online violence against women journalists.

Platforms could:

- Continuously review their policies, algorithms and moderation processes, to address the ever-evolving nature of online violence.

- Implement proactive countermeasures against abuse and reverse the onus on the women victims to report the abuse to start with.

- Define more effective policies and procedures for detecting and penalising repeat offenders, and stop the same abusers assuming new online identities after temporary suspensions or permanent bans.

- Create more effective abuse reporting and content moderation tools that support local languages and are sensitive to local cultural norms, because harassment and abuse received by women journalists is often posted in a minority language, dialect, vernacular or slang.

- Develop reporting systems with capacity for escalation for women journalists under attack (and their employers), recognising their particular vulnerabilities and the implications for press freedom, and include the problem and reactions in regular transparency disclosures.

Media organisations could:

- Introduce or update protocols and guidelines pertaining to online violence to ensure they are gender-sensitive and gender-responsive, and develop appropriate responses in the context of weaponised social media platforms, viral disinformation, far-right extremism and conspiracy networks.

- Hold the platforms to account through critical reporting, and advocacy on media freedom and journalism safety, regardless of commercial ties to the social media companies.

- Ensure online safety support is holistic (integrating psychological, digital security, editorial, and legal responses), as well as responsive to intersectional threats/impacts, and readily available to all staff and freelancers.

- Make it clear that it is not appropriate for staff to participate in acts of targeted online violence (including trolling) against women colleagues or those working for competing news outlets.

Civil society organisations and donors could:

- Reinforce the call for responses to online abuse of women journalists to conform to international human rights standards.

- Partner with journalists, news organisations and researchers on investigative and monitoring projects about how online violence manifests itself, and responses to it.

- Help raise awareness and educate women journalists and editors in online safety, effective use of platform tools for countering online abuse, employer advocacy, and legal support.

A note about our methodologies

The survey method adopted was 'purposive sampling', with 'snowballing' techniques used to generate responses within the international field of journalism. The results, therefore, are not generalisable, although it is legitimate to extrapolate many patterns that may well have wider applicability. To avoid illegitimate or inauthentic responses and ensure data integrity, the survey was distributed digitally via the closed networks of UNESCO and ICFJ, our research partners, civil society organisations focused on media development, journalism safety and gender equality, and groups of professional journalists. The survey ran from September 24th to November 13th 2020 and it garnered 901 valid responses. The survey results were then disaggregated along gender lines, and a subset of data from 714 respondents who identified as women was isolated for analysis. In parallel, we identified 173 interviewees through the survey and institutional outreach, as well as via the networks of the research team. The interviews were conducted face-to-face (where COVID-19 restrictions allowed) and via digital channels. Most of the interviews were undertaken synchronously by the researchers identified in this report. The vast bulk of interviewees chose to be publicly identified after being offered the option to remain anonymous.

For the big data case studies on Maria Ressa and Carole Cadwalladr 2.5 million social media posts were collected over the course of five years and 13 months respectively. Relevant subsets of these collections were identified for network analysis and deeper investigation via Natural Language Processing (NLP). The results were synthesised with the long form qualitative interviews and contextualised via detailed timelines developed through desk research.

The University of Sheffield (UK) granted ethics clearance for the English language version of the survey and English language interviews. Translations of the survey into other languages were conducted by UNESCO and reviewed by ICFJ. The University of Sheffield also provided ethics clearance for quantitative data gathering and analysis associated with the big data case studies featured here.



in cooperation with

