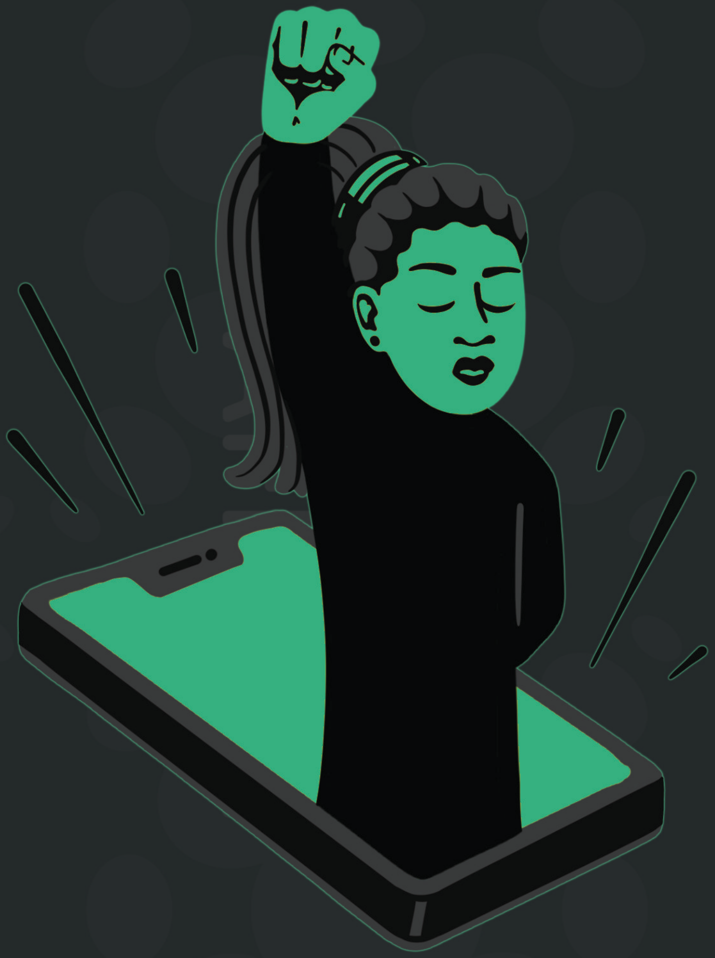

SAFEGUARDING THE FRONTLINES

a digital security toolkit for activists and human rights defenders

MEDIAMONITORING AFRICA

POWERLAW /AFRICA

CONTENTS

SAFEGUARDING THE FRONTLINES
A digital security toolkit for activists and human rights defenders

This toolkit aims to equip activists, human rights defenders, and public interest lawyers with the tools, knowledge, and support they need to guard against, navigate and respond to online harassment and abuse.

Published by Media Monitoring Africa, 2025.

Drafted by Sophie Smit and Christy Chitengu, Power Law Africa.
Design and typesetting by ALT Media.

Cover illustration: Johnny Allison / ALT Advisory (CC BY-NC-ND 4.0)

INTRODUCTION

Human Rights Defenders (HRDs) play a crucial role in exposing injustice, protecting vulnerable populations, holding power to account, and defending fundamental freedoms. HRDs are an essential component of not just protecting democracies but breathing life into them and keeping them alive. Yet, those who speak out, particularly online, often become targets of coordinated attacks, harassment, intimidation, and suppression. Digital spaces, while offering vital tools for advocacy, also expose HRDs to serious risks and threats.

It is important to note upfront that digital threats are diverse. Online harassers use many tools and tactics, including social media trolling, hateful speech, direct threats, doxxing (publishing personal information), hacking, sharing or misuse of images without consent. These digital threats may even escalate to legal harassment which may include lawsuits and/or arrests, online surveillance and smear campaigns.

WHAT THIS TOOLKIT AIMS TO DO

This toolkit aims to prepare HRDs for safer digital engagement by following the acronym **I-PREP**:

I-PREP

Identify online risks specific to your context.

Prevent online harassment with practical strategies.

Respond to online harassment through support mechanisms.

Empower resilience and collective support.

Promote advocacy for systemic change.

By offering concrete tools and knowledge, we hope this toolkit will empower you to reclaim safer, more resilient digital spaces and continue vital human rights work with greater security and confidence.

WHY THIS MATTERS

PREVALENCE AND IMPACT OF ONLINE HARASSMENT AGAINST HUMAN RIGHTS DEFENDERS

The scale of harassment against human rights defenders in digital spaces is large and rising. Recent events in the global north have emboldened misogynists and racists, and equally disturbingly, also served to highlight polarisation. The decline in trust toward credible sources, coupled with digital media platforms that reward sensationalism and the ease of anonymously mobilising troll armies, has made attacks on HRDs not only more visible but increasingly normalised.

Below are some of the key findings and statistics that show how urgent and pervasive the problem is:

- **Young human rights defenders are highly exposed.**

A global survey by Amnesty International of over 1,400 young activists (ages 13-24) in 59 countries found that "three out of five Child and Young Human Rights Defenders face online harassment in connection with their activism."[1] Common forms of harassment included hateful comments, threats, hacking and doxxing. In addition:

> "Twenty-one percent of respondents say they are trolled or threatened on a weekly basis and close to a third of the young activists say that they have censored themselves in response to tech-facilitated violence, with a further 14 percent saying they have stopped posting about human rights and their activism altogether."[2]

- **Environmental and land defenders are nearly universally affected.**

A survey conducted by Global Witness during November 2024 to March 2025 found that Land and Environmental Defenders often rely on digital platforms to organise, share information and campaign.[3] According to survey findings, 92% of these defenders said they experienced some form of online abuse or harassment as a result of their work.[4] The types of abuse included cyberattacks, doxxing, harassment of images, attacks on character and "frequently translates into offline harm, including harassment, violence and arrests."[5]

- **Woman Human Rights Defenders are particularly at risk**

According to the United Nations Secretary-General, anti-rights actors are increasingly using online platforms to push back against women's rights, targeting Women Human Rights Defenders with tactics that include cyberbullying, harassment, threats of violence, gendered disinformation, and threats.[6] A survey of 458 Women Human Rights Defenders conducted by the Kvinna till Kvinna Foundation in 2023 found that 75% said that either they or their organisation had faced threats or harassment over their work and that almost 1 in 4 had received death threats.[7]

- **Woman journalists are increasingly being targeted online for their work**

Journalists are facing increasingly severe digital attacks designed to discredit and intimidate them, with women journalists among the hardest hit. In a global UNESCO-ICFJ survey of 901 women journalists across 125 countries, 73% reported experiencing online violence, including harassment, threats, insults and attacks tied to their work.[8] About 25% had received threats of physical violence, and 18% had been threatened with sexual violence.[9] Alarmingly, 20% said online abuse had escalated into offline attacks or harassment. Another survey of more than 400 women journalists by Women in Journalism/Reach in the UK in 2023 found 75% had experienced threats or safety challenges (online or offline) in the course of their work, and many reported scaling back their online presence or considering leaving journalism altogether.[10]

[1] Amnesty International "Three out five young activists face online harassment globally for posting human rights content" (1 July 2024). (Accessible here.)
[2] Id.
[3] Global Witness "Toxic platforms, broken planet: How online abuse of land and environmental defenders harms climate action" (16 July 2025). (Accessible here.)
[4] Id.
[5] Id.
[6] UN Women "How women human rights defenders are under threat worldwide" (26 November 2024). (Accessible here.)
[7] The Kvinna till Kvinna Foundation "Hope and Resistance Go Together: The State of Women Human Rights Defenders 2023" (14 December 2023). (Accessible here.)
[8] UNESCO "UNESCO's Global Survey on Online Violence against Women Journalists" (15 December 2020). (Accessible here.)
[9] Id.
[10] Reach "Survey reveals 75% of women journalists have experienced a threat to their safety" (8 March 2023). (Accessible here.)

AN OVERVIEW OF THE NUMBERS

- Three out of five Child and Young Human Rights Defenders face online harassment in connection with their activism.
- 92% of Land and Environmental Defenders said they experienced some form of online abuse or harassment as a result of their work.
- 75% of Woman Human Rights Defenders said that either they or their organisation had faced threats or harassment over their work and that almost one in four had received death threats.
- 73% of women journalists reported experiencing online violence, including harassment, threats, insults and attacks tied to their work.

TAKING ACTION

I-PREP

1: IDENTIFY RISKS

Why it matters: You can't defend yourself if you don't know your vulnerabilities. Online harassment often exploits the information HRDs leave exposed.

The first step in staying safe online is to identify risks. This means mapping your digital footprint by searching for your name, usernames, or images online and reviewing what information appears publicly. It also involves auditing your accounts to check privacy settings and the visibility of your posts and connections. Tools created by the Information Regulator and New York Times can help you assess your privacy settings and follow good practice. By doing this, you can see what personal details might be misused by harassers. It is equally important to understand who would be most likely to target you or your organisation, whether they are individuals, organised groups, or state actors, and to take note of past patterns of harassment.

A simple digital risk log (see Annexure A), recording what is exposed, the level of risk, and the action you have taken, can be an effective way to track and reduce vulnerabilities.

2: PREVENT HARASSMENT THROUGH PROACTIVE DIGITAL PROTECTION

Proactive digital protection acts as a safeguard for activists and HRDs. In an increasingly digitised world, individuals must take proactive steps to protect their digital security. Curating your digital presence allows you to maintain a strategic digital footprint which is crucial for activists and HRDs. A well-managed online presence balances visibility with privacy, ensuring that personal information does not become an avenue for attacks, harassment, or surveillance.

Activists and HRDs may consider creating separate accounts for activism and personal use as an effective way to protect personal privacy. If this is something that an actor in this space would consider, then public accounts used for advocacy should be professional, well-curated, and devoid of personally identifiable information, such as home addresses, phone numbers, or private email addresses. Using aliases or pseudonyms for activism-related accounts can also enhance anonymity. Many activists choose to operate under different names to minimise the risk of being targeted.

Whether these digital spaces include social media accounts, email or other digital spaces, by implementing robust security measures before threats arise, activists and HRDs can reduce vulnerabilities and maintain a strong digital presence while ensuring their safety.

Fortify Your Access

The foundation of digital security begins with securing access to online accounts and devices. Weak passwords and a lack of multi-layered authentication mechanisms can make individuals vulnerable to cyberattacks, identity theft, and unauthorised access. The following steps should be taken to ensure that access to your online accounts and devices are secure:

1. **Ensure that you have strong and unique passwords:** Passwords serve as the first line of defence against cyber intrusions. Weak, easily guessable passwords are one of the primary ways hackers gain access to sensitive accounts. A strong password should be at least 12–16 characters long and include a mix of uppercase and lowercase letters, numbers, and special characters. Using a password manager can help generate and store complex passwords securely. Many password managers, such as Bitwarden, LastPass, or 1Password, allow users to create unique passwords for each account without the need to memorise them. This practice prevents attackers from exploiting reused passwords across multiple platforms.

2. **Two-Factor Authentication for Enhanced Security**: While strong passwords are essential, they are not enough on their own. Two-factor authentication ("2FA") adds an extra layer of security by requiring users to verify their identity using two different forms of authentication. This often includes something the user knows (password) and something the user has (a mobile device or security key). Enabling 2FA on social media, email, banking, and advocacy-related platforms significantly reduces the risk of unauthorised access. Authenticator apps, such as Google Authenticator or Authy, are more secure than SMS-based 2FA, as text messages can be intercepted by attackers. Hardware security keys, like YubiKey, provide even greater protection against phishing attacks.

3. **Routine Security Checks:** Regularly monitoring digital accounts helps detect potential breaches before they cause harm. Reviewing login activity on platforms like Gmail, Facebook, and Twitter can help identify suspicious access attempts. If any unusual activity is detected, immediately change passwords and revoke unauthorised access.

4. **Adjusting Privacy Settings:** Social media platforms provide various privacy settings that allow users to control who can view their posts, send friend requests, and interact with their content. Activists should regularly review and update these settings to restrict public access where necessary. For instance, on Facebook, enabling profile lockdown settings prevents non-friends from viewing posts and accessing contact details. On Twitter/X, using protected tweets limits visibility to approved followers only. On LinkedIn, adjusting search visibility ensures that personal information is not easily accessible to unwanted individuals.

5. **Minimising Digital Traces:** Reducing personal exposure online involves taking steps to remove or obscure sensitive information. People-search websites and data brokers collect personal information from public records, often without

consent. Services like DeleteMe can help remove personal details from these databases. Additionally, regularly searching one's own name on Google and requesting takedowns of sensitive information can minimise digital traces.

6. **Shield Your Devices:** While securing accounts is essential, protecting the physical and digital integrity of devices is equally important. Smartphones, laptops, and other digital tools can be compromised by malware, hacking attempts, or physical confiscation. Taking proactive measures ensures that sensitive data remains secure.

7. **Protecting Against Cyber and Physical Threats:** Activists often work in environments where both cyber and physical security are at risk. To protect against unauthorised access, consider the following measures:

   - Enable full-disk encryption on laptops and smartphones to prevent data theft if a device is lost or confiscated.
   - Use biometric authentication (fingerprint or facial recognition) alongside strong passwords for device security.
   - Keep software updated to ensure the latest security patches are applied, reducing vulnerabilities that attackers could exploit.

8. **The Role of Virtual Private Networks ("VPNs") in Digital Security:** When working online, especially in high-risk environments, it's essential to be mindful of the networks you connect to. Public WiFi networks, such as those in cafés, airports, or hotels, are often insecure and can expose sensitive communications to interception. Whenever possible, reserve sensitive work for more secure connections, such as your home or trusted organizational networks. VPNs can offer an added layer of protection by encrypting your internet traffic and masking your IP address. This helps prevent internet service providers, government surveillance, and malicious actors from tracking your online activity, particularly useful when accessing the internet over untrusted networks. However, it's important to note that not all VPNs are created equal. Because a VPN routes all your network traffic through its servers, you must place a high degree of trust in the provider. Some VPNs have been found to log user data, inject ads, or even sell information to third parties. Before using a VPN, it's crucial to research and select a reputable provider with a strong privacy track record. Consulting reliable sources such as Wirecutter's VPN reviews can help users make informed decisions.

I-PREP

3: RESPOND TO HARASSMENT

Why it matters: Harassment escalates if unmanaged. A clear response plan prevents panic and protects you legally and digitally.

Responding promptly and effectively to a digital threat is essential to protecting yourself and mitigating further harm. This section outlines key pathways for action, ranging from blocking and reporting harmful content to accessing legal remedies and leveraging support. Whether the threat arises from disinformation, targeted harassment, or hate speech, a well-informed response can make a meaningful difference.

The first step after encountering harmful content online is to ensure that you record it by either screen grabbing or copying the URLs. This ensures that there is evidence, and a public record of the harmful content should it be necessary. After it is recorded, one may use two specialised mechanisms developed by MMA to report the harmful content: Real411[11] and the Media Attack Reporting System (MARS).[12]

MARS is a dedicated platform for documenting threats against journalists. During the 2024 election period alone, MARS recorded over 1,025 incidents of online abuse, with a significant number targeting women journalists. The platform plays an important role in identifying patterns of abuse and supporting press freedom.

Real411, launched in 2019 by MMA in collaboration with the Electoral Commission (IEC), is a public platform that allows individuals to report four categories of online harms: disinformation, harassment, hate speech, and incitement to violence. Originally designed to support the integrity of electoral processes, Real411 has grown into a year-round accountability tool. Complaints are assessed by a trained panel of experts, including legal professionals, academics, and IT specialists, and decisions can be appealed. The platform is transparent, with a publicly accessible archive of complaints, and serves as an early warning system by tracking trends in online abuse.

[11] Real 411. (Accessible here.)
[12] Media Attack Reporting System. (Accessible here.)

Notably, retired South African Constitutional Court Justice Zak Yacoob oversees the appeals process, adding credibility and trust to the mechanism.

After the content has been reported to Real411 and MARS, a user can use the reporting and blocking functions provided by social media platforms. On X (formerly Twitter), users can block individuals and report abusive tweets directly through the platform's interface. Similarly, Instagram, Facebook, TikTok, and LinkedIn all offer in-app options to report offensive content and block users. These tools help to immediately limit further exposure and flag inappropriate behaviour for the platforms to review. However, platform responses may vary, and the definitions of harassment or harmful content used by platforms often differ from those defined under South African law.

In addition to these reporting tools, South African law provides a vital legal remedy through protection orders, as outlined in the Protection from Harassment Act.[13] Victims of online harassment can apply for a protection order, which legally prohibits the perpetrator from continuing their behaviour. If breached, the offender may face fines or imprisonment.[14] When the identity or physical address of the harasser is unknown, the courts may compel electronic communications service providers, including social media platforms, to release identifying information. However, it is important to note that what qualifies as harassment under the Act may not meet the definition used by platforms. This discrepancy can make it challenging for victims to get harmful content removed unless they pursue costly and time-consuming legal processes.

Responding to online harms involves a multi-pronged approach that combines technical action, formal reporting mechanisms, legal remedies, and community support. By understanding and accessing these different pathways, individuals and organisations can protect themselves more effectively and contribute to a safer, more accountable digital environment.

Emergency Quick-Response Guide

1. **Stop engaging** – don't reply to harassers.
2. **Document evidence** – screenshot any posts or messages and save links.
3. **Report the harmful content** – to Real411[15] and the Media Attack Reporting System (MARS).[16]
4. **Inform a trusted contact** – never go through it alone.
5. **Report to platforms** – and escalate if ignored.
6. **Assess physical security** – update your routines if threats escalate offline.

[13] The Protection Against Harassment Act 17 of 2011 (accessible here) defines harassment to include online harassment, as seen at section 1.
[14] Id at section2.
[15] Real 411. (Accessible here.)
[16] Media Attack Reporting System. (Accessible here.)

I-PREP

4: EMPOWER RESILIENCE AND COLLECTIVE SUPPORT

Why it matters: Harassment is designed to isolate you. Solidarity and mental resilience are your strongest shields.

Harassment is designed to isolate and silence defenders, which is why building resilience and support networks is so important. It is critical to engage and inform other activist and HRD organisations when online harms occur. Notifying trusted networks, such as digital rights groups, journalist collectives, or legal advocacy organisations, can help amplify awareness of abusive trends, facilitate coordinated responses, and provide victims with emotional and legal support. Collective action not only strengthens individual resilience but also builds solidarity and shifts the broader culture around online abuse.

Establishing support circles ensures that defenders do not face harassment alone. Simple practices such as weekly check-ins, team-wide digital literacy training, and shared security practices strengthen collective resilience.

Mental health care is also part of digital safety, whether through staff debriefs, professional counselling, or setting healthy boundaries with technology. A short staff-support check can make a significant difference during periods of intense harassment.

I-PREP

5: PROMOTE SYSTEMIC CHANGE

Why it matters: Individual security is vital, but lasting protection comes from stronger systems: accountable platforms, supportive laws, and global solidarity.

While individual precautions are crucial, long-term protection depends on systemic change. Activists and defenders can contribute by supporting digital rights campaigns, joining coalitions that advocate for stronger platform accountability, and calling for better legal protections. Sharing experiences of harassment, when it is safe to do so, can highlight patterns and strengthen advocacy efforts. Even anonymised stories help build evidence and put pressure on platforms and governments to act. Creating an advocacy action plan with clear goals, identified allies, and defined tools, whether petitions, reports, or policy briefings, can help defenders engage in systemic change while protecting themselves.

CONCLUSION

Online harassment poses a significant and growing challenge for HRDs and activists worldwide. It can take many forms, from threats and trolling to surveillance, doxxing, and coordinated smear campaigns, and its impacts extend beyond the digital space, affecting personal safety, mental health, and the ability to continue vital work.

By systematically assessing digital risks, implementing preventive measures, responding strategically to incidents, fostering support, and engaging in advocacy for safer digital environments, defenders can strengthen both individual and collective security. While no strategy can eliminate all risks, preparation, awareness, and solidarity dramatically increase safety and resilience.

Ultimately, this toolkit is more than a set of technical instructions, it is a guide for sustaining human rights work in a digital age. By using these strategies, defenders can continue their advocacy with confidence, knowing that they are equipped to navigate online threats and protect themselves, their communities, and their mission.

USEFUL RESOURCES:

Media Defence (Prepared by ALT Advisory)

Practical Approaches to Combatting Online Violence Against Women Journalists

Media Defence's Summary Modules on Digital Rights and Freedom of Expression consist of ten modules designed as a reference guide for litigating cases of digital rights in sub-Saharan Africa. These modules are aimed at an audience of lawyers, with experience of litigation, but not necessarily of media, digital rights, freedom of expression or human rights law. They can also be useful to law students and legal practitioners with an interest in digital rights.

Access Now:

Access Now – Digital Security Helpline (24/7)

This is a free-of-charge resource for civil society around the world. It offers real-time, direct technical assistance and advice to civil society groups and activists, media organisations, journalists & bloggers, and human rights defenders.

Front Line Defenders:

Front Line Defenders – Digital Protection Resources / Digital Protection Kit

The Front Line Defenders Digital Protection programme responds to the digital security environment facing HRDs and develops tools, guides and resources to complement its training and consultation programming.

Amnesty International

Amnesty International – Security Lab Resource Hub

The Digital Security Resource Hub was prepared by Amnesty International's Security Lab for human rights defenders, activists, journalists and other members of civil society. This hub provides an overview of a variety of resources ranging from risk analysis support to helplines to tools that you can use to strengthen your digital and information security practices.
)

Kvinna till Kvinna / Resources for Women & LGBTQ+ Defenders

Kvinna till Kvinna: digital learning & publications

The State of Women Human Rights Defenders 2023

Thomson Reuters Foundation

Weaponizing the Law: Attacks on Media Freedom

Annexure: Digital Risk Log Template

This Digital Risk Log is provided as a tool for Human Rights Defenders (HRDs) and activists to monitor, assess, and manage online risks. It should be completed regularly as part of digital security practices and may be adapted for individual or organisational use.

Digital Risk Log

Date	Platform / Account	Type of Data Exposed	How Was It Found?	Risk Level (Low / Medium / High)	Action Taken	Follow-Up Required

Guidance Notes

Date: Enter the date when the risk was identified.

Platform / Account: Indicate the platform (e.g., Facebook, Twitter, LinkedIn) and specify the account.

Type of Data Exposed: Note whether it is personal information, contact details, workplace data, images, or geolocation.

How Was It Found: Record whether the exposure was identified through self-search, peer review, monitoring tool, or incident report.

Risk Level: Assess whether the exposure poses a *Low*, *Medium*, or *High* risk.

Action Taken: Describe the immediate steps taken to reduce or eliminate the risk (e.g., removed, adjusted privacy settings, deleted account).

Follow-Up Required: Note any further steps, such as re-checking cached data, monitoring reposts or conducting a periodic re-assessment.