Results overview:
Pilot study on experiences with social media and communication platforms in MENA and East Africa regions

June 2023

UNITED NATIONS
HUMAN RIGHTS
OFFICE OF THE HIGH COMMISSIONER

Pilot study aims and scope

The present pilot study provides an overview of the experiences of Human Rights Defenders (HRDs) and journalists in the Middle East, North Africa (MENA) and East Africa[i] who use social media and communication platforms in their professional capacity.

OHCHR conducted this study with the aim of developing tools to improve its understanding of civil society's experiences in the digital space and informing ongoing debates about the ways social media and communication platforms respond to civil society in the Global South, paying special attention to protection demands from HRDs and journalists.

HRDs and journalists are central civic space actors who promote public debate and participation. Given their role, they commonly attract public attention and are therefore likely to become targets of different forms of online threats and attacks, often via social media and communication platforms. Collecting information about their experiences and listening to their concerns is key to understanding how these new communication channels are transforming civic space and what may be done to ensure they provide reliable, accessible, and safe lines of communications for all.

The pilot study focuses on the experiences of HRDs and journalists who work in MENA and East Africa, as both regions experience a range of online civic space concerns, such as the expansion of State capacities to monitor online spaces and the enforcement of a wide range of legal instruments that further limit online freedom and privacy. Parts of the regions also encounter specific risks posed by armed conflict, where communications are critical for safety as well as for documenting rights abuses.

The data presented in this document summarizes some of the findings of a pilot survey[ii] circulated between November 2022 and February 2023. The survey had 315 respondents where 47% identified themselves as HRDS, 36% as journalists, and 17% as "other civil society actors", with a gender balance of 56% male, 43% female, and 1% other/non-binary. There was also geographic balance, with 54% of respondents from MENA and 46% from East Africa, covering 27 countries in total. Most respondents (89%) were based in their home country, while others work on the situation of countries in the region while based abroad (often in exile).

The illustrative testimonies presented here were extracted from semi-structured interviews conducted between January and April 2023. A total of 75 individuals were interviewed in this process, with 34 respondents taking part in six group discussions (3 in each region) and the rest participating in individual interviews. Additional 24 civil society organizations working on human rights and media freedom issues in the regions contributed with dissemination of the survey, identification of interlocutors, and expert input.

OHCHR is very grateful for the collaboration of all individuals who shared their experiences and for the contributions of civil society organizations to this pilot study, in particular Access Now, which provided direct support for the translation of the survey, in addition to valuable expertise and contacts.

Platforms in use

All HRDs and journalists interviewed reported heavily relying on social media and messaging platforms, which are regarded as key tools for research, networking, outreach, and disseminating information. According to the online survey, the most used platforms are under Meta (Facebook, WhatsApp, Messenger, and Instagram), with 51% of responses. Other commonly used platforms include Twitter, Telegram, YouTube, LinkedIn, Signal, and TikTok.

Select the social media and messaging platforms you currently use:
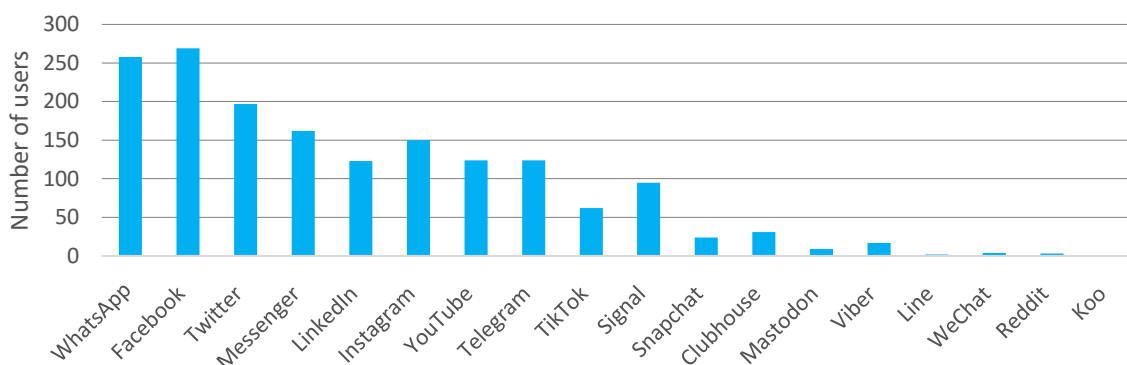
About 37% of total respondents (mostly HRDS) stated that they have used free or paid tools to promote the visibility of their publications or profiles on social media platforms. Most were satisfied with the resulting visibility, yet cost barriers were raised as a concern.

Have you ever used tools to promote content visibility?

- Yes, and I am satisfied
- Yes, but I am not satisfied
- No
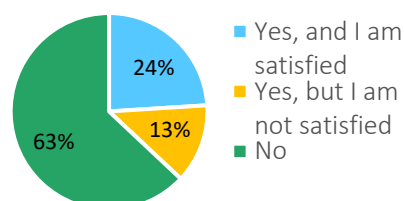
Experiences of threats and attacks

A majority of HRDs and journalists surveyed reported having experienced some form of threat or attack. Almost half experienced attacks and threats online. Interviewees who experienced attacks online and offline often indicated receiving online messages as "warnings" before other offline events.
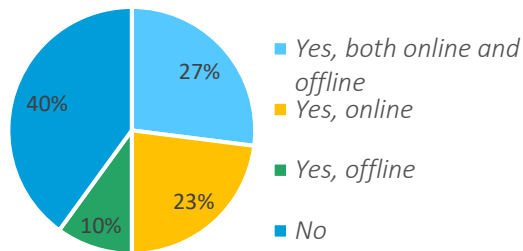
The most common forms of online threats or attacks reported were smear or defamatory campaigns, followed by direct threats of violence; account hacking or phishing;[iii] and false impersonaiton[iv] and doxing.[v] Among respondents who experienced online attacks, 48% reported experiencing one of the categories of threats or attacks, while 52% reported experiencing multiple threats.

"I've had some co-workers who've been threatened for some controversial social media posts, which has made us all worried about our safety. When we post something, we might get phone calls, online threats in social media, or even death threats, sometimes through direct calls. In my country, kidnapping is a real issue, and it's not hard to find yourself in a dangerous situation. A few of my colleagues who work in advocacy began receiving threats online, but these eventually turned into real-life threats. They reported it to the police, but unfortunately, nothing positive has come from it."
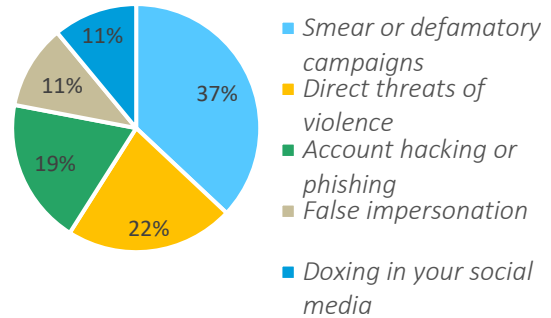
TESTIMONY - FEMALE NGO WORKER AND HRD (EAST AFRICA)

When asked to identify the reasons behind these attacks, most respondents related the attacks to their activities as HRDs and journalists, as well as the ideological or political affiliations of their profiles. The majority of HRDs and journalists who reported threats and attacks said they experienced multiple incidents.
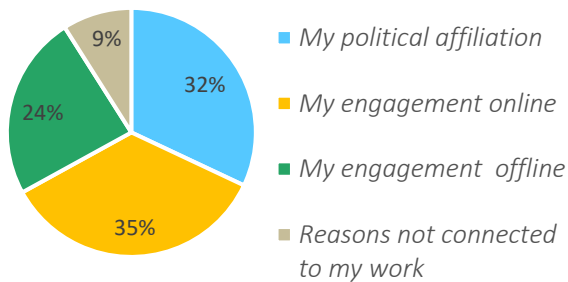
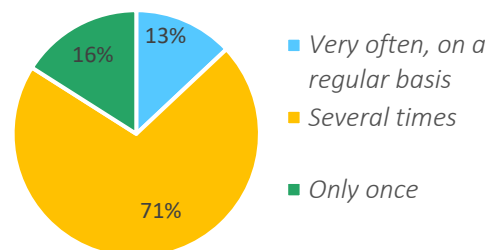*Have you ever been the target of online or offline threats?*

- Yes, both online and offline — 27%
- Yes, online — 23%
- Yes, offline — 10%
- No — 40%

Select the online threats or attacks you have been the target of:

- Smear or defamatory campaigns — 37%
- Direct threats of violence — 22%
- Account hacking or phishing — 19%
- False impersonation — 11%
- Doxing in your social media — 11%

What do you think are the main reason(s) behind those threat(s) or attack(s)?

- My political affiliation — 32%
- My engagement online — 35%
- My engagement offline — 24%
- Reasons not connected to my work — 9%

How often have you experienced the threats online?

- Very often, on a regular basis — 13%
- Several times — 71%
- Only once — 16%

HRDs and journalists who self-identified as female, non-binary, and male in the survey reported similar prevalence in the categories of threats and attacks they experienced. However, interviews with HRDs and journalists, as well as organisations that support women HRDs and LGBTQI+ communities, reported a gendered dimension to the threats and attacks experienced by these groups. Several interviewees cited harassment and intimidation with elements of sexual character; threats of a sexual character, including rape threats; smear and defamation campaigns based on sexualization, in some cases using pictures of them (including fake images) that damaged their reputation; and death threats.

Interviewees of all genders raised organized and coordinated threats and attacks as important factors in silencing dissenting voices. Coordinated campaigns involving multiple forms of attacks were often attributed to groups defending Governments' interest in silencing critical voices, both within borders as well as overseas. Common tactics included hiring of social media influencers to spread disinformation or directly harass and discredit civic space actors, as well as trolls and automated messaging systems to amplify defamatory or hostile messages.

Did you establish contact with the platform in connection with the threats or attacks?

- Yes, and I obtained a response
- Yes, but I did not obtain a response
- No

11%
13%
75%
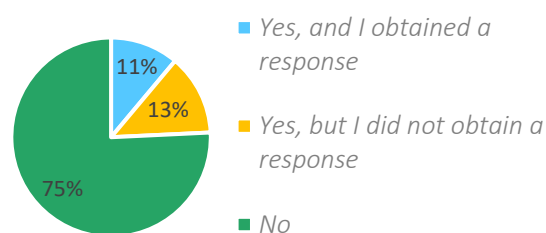
The majority of surveyed HRDs and journalists threatened or attacked online said they did not report these incidents to the platforms. Among those who contacted the platforms, a slight majority obtained responses.

> *"In my country, specialized police departments are targeting journalists and dissenting voices in general, and they use social media to conduct defamation and smear campaigns. Often, they use what is considered red lines in our society, especially sexuality, honour… Many journalists have been victims of smear campaigns in social media, often with false stories and images fabricated and disseminated. Women journalists and HRDs are especially vulnerable to those attacks."*
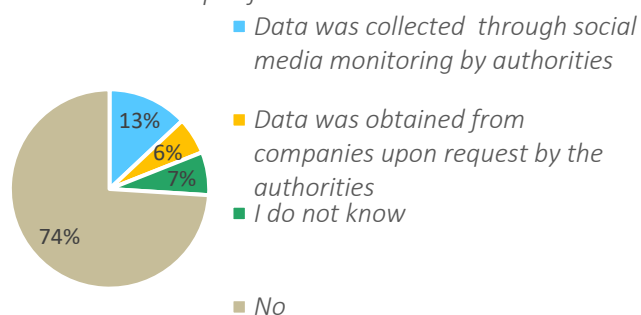>
> TESTIMONY – FEMALE JOURNALISTS (MENA)

Information collection and legal repercussions of online posts

Threats and attacks were in some cases associated with the collection of information from social media accounts. About a quarter of HRDs and journalists report believing that that law enforcement or other authorities collected information about them from their social media accounts. Among that group, half believed data was collected through the direct monitoring of accounts, while others believed data was collected from companies or were unsure of the means used to extract data.

The extraction of private information was mentioned as a common component of some attacks targeting women and LGBTQI communities. Some interviews underlined that the mere exposure of women's activism can, in some contexts, be perceived as contravening traditional values and be used to trigger smear campaigns.

Have you ever been subject to authorities collecting private data from social media and communication platforms?

- Data was collected through social media monitoring by authorities
- Data was obtained from companies upon request by the authorities
- I do not know
- No

13%
6%
7%
74%

In both regions, interviewees raised concerns regarding the adoption of new laws that expand restrictions for online expression. In accordance with the evolving national legal frameworks, HRDs and journalists faced legal charges in relation to social media activities, including for allegedly disseminating hate speech, spreading false information, inciting violence, attacking state authorities,

tarnishing the country's image, endangering national security, and violating traditional and family values.

Almost a quarter of survey respondents reported having been the target of legal measures that used online communication as evidence for charges, with a great chilling effect on other HRDs and journalists. In connection with potential risks represented by authorities' access to content, 18% of respondents requested the deletion of their social media or messaging platform accounts or digital footprints at least once. Being able to rapidly lock and hide human rights defenders' and journalists' accounts was raised as a potential shield to deter both further legal measures against them and as a means of preventing potential repercussions for professional associates who could also be exposed if authorities accessed their communications data. Among those who requested the deletion of their account, 54% reported that the request was successful.

Have you been subject to legal actions using online communication as evidence?

22%

78%

■ *Yes* ■ *No*
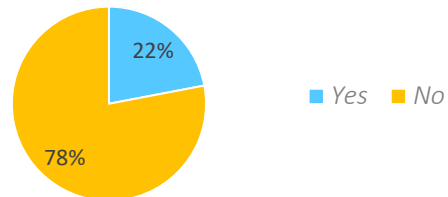
"When HRDs are vocal in our communities, they become targets, and their social media accounts are monitored, their communications are mostly wiretapped, and put under surveillance The police harass and intimidate us both online and offline. Some of us have even been arrested because of what we've posted online. The information we share on social media can be used against us in court. Sometimes the authorities ask us to remove certain posts, and if we don't, we might receive phone calls and threats, either online or offline. You must think twice before you post on sensitive issues, otherwise, you might get in troubles."

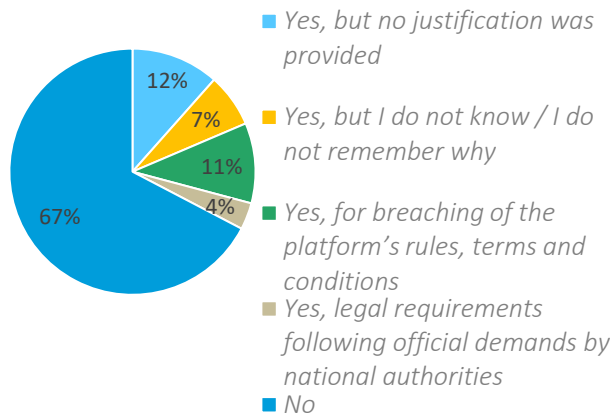TESTIMONY - GROUP OF HRDS AND JOURNALISTS (EAST AFRICA)

Experiences with content moderation

The pilot study also collected information on HRDs' and journalists' experiences with the enforcement of platforms' content moderation policies. Awareness of existing standards and channels to contact social media platforms varied greatly in accordance with respondents' frequency of use of platforms and contacts with specialized networks. The most common and important concern was the blocking of accounts or of specific content.

Around a third of the surveyed HRDs and journalists reported having their profiles or accounts blocked or suspended by the platforms. Among those who reported being affected by blocking, the majority indicated that no justification was provided or that they did not recall the reasons explaining the blocking or suspension, followed by those who said their blocking was justified by alleged breaches of platform rules. In a minority of cases, the blockings were allegedly imposed in response to legal requests posed by authorities.

The majority of respondents whose accounts were blocked requested the restoration of their accounts; around half of those requesting managed to have their profiles restored.

Has any of your profiles or accounts ever been blocked? Were you provided with a justification?

- Yes, but no justification was provided
- Yes, but I do not know / I do not remember why
- Yes, for breaching of the platform's rules, terms and conditions
- Yes, legal requirements following official demands by national authorities
- No

12%
7%
11%
4%
67%

Did you ask the social media platform to restore your account or profile? And was your content restored in the end?

- Yes, and it was restored
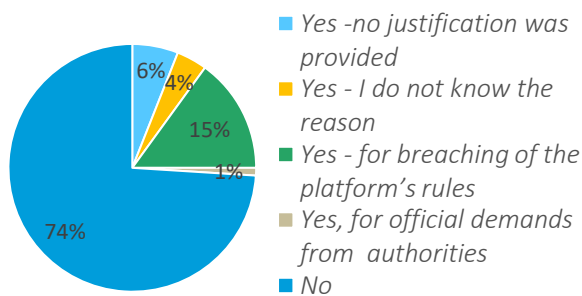- Yes, but it was not restored
- No

19%
45%
36%
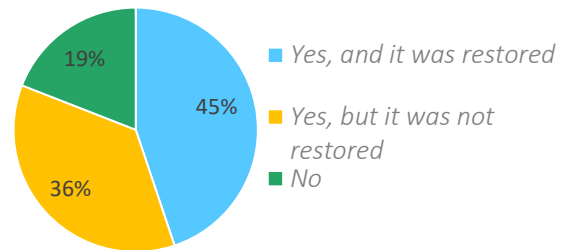
"A few years ago, more than 12 social media accounts of journalists from our country were deleted upon government request, claiming it was to counter propaganda. This was well documented by human rights organizations."

TESTIMONY - GROUP OF JOURNALISTS (EAST AFRICA)

Similarly, around a third of the surveyed HRDs and journalists reported having elements of published content deleted. The alleged breach of platform's rules, terms, and conditions were provided as the main reasons for the imposition of specific content bans.

Most surveyed participants also requested reinstating the blocked content, but the majority of those requests were unsuccessful. Account and content blocking concerns were particularly prevalent among participants in the MENA region.

Have any of your posts been banned or deleted by social media platforms? Were you provided with a justification?

- Yes -no justification was provided
- Yes - I do not know the reason
- Yes - for breaching of the platform's rules
- Yes, for official demands from authorities
- No

6%
4%
15%
1%
74%

Did you ask the social media platform to restore your posts, content, or messages? What was the result?

- Yes, and it was restored
- Yes, but it was not restored
- No

23%
38%
39%
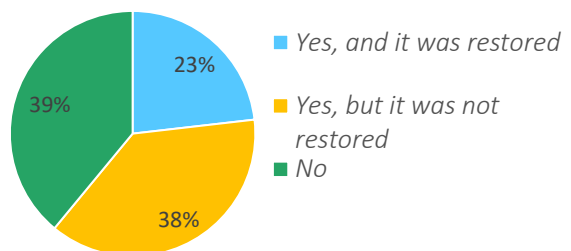
Have you faced disruptions of livestreamig unrelated with connectivity issues?
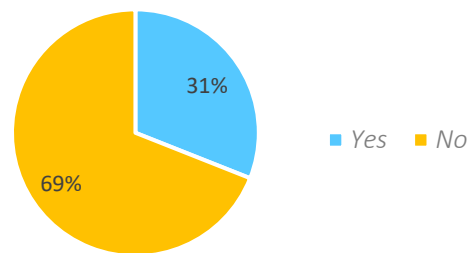
24%

76%

■ *Yes* ■ *No*
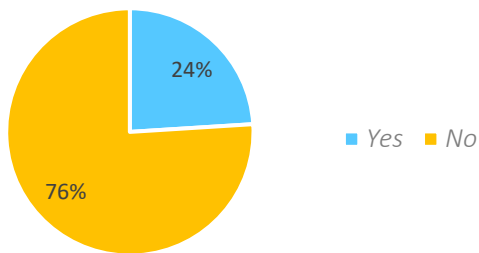
Have you experienced shadow banning or demotion in publications?

31%

69%

■ *Yes* ■ *No*

About one quarter of respondents from both regions reported facing operational disruptions during livestreams or communication on social media platforms unrelated to device connectivity. Following the same pattern, approximately one-third of the survey respondents indicated that they believe to be affected by some form of "shadow banning," suspecting that their publications might have been demoted or downgraded, thereby reducing their exposure and reach.

Interviewees also noted the abuse of platforms' channels for flagging problematic content, specifically through mass reporting, which is a common tactic exploited by Governments and other powerful actors to undermine or inhibit communication. As a result of the recurrent risk of account removal or suspension and the perceived unpredictable processes of reestablishment, some HRDs and journalists interviewed keep multiple spare accounts (mainly Facebook and Twitter were mentioned) as a mitigation tool.

"Organized trolls are a reality in our country and usually the government is behind. They are typically coordinated through WhatsApp or Telegram groups to launch massive attacks to specific social media profiles. Once those profiles are massively reported by the organized group, they are automatically blocked, and it usually takes long for the victim to restore the account, if successful. Journalists are common victims of coordinated attacks."

TESTIMONY - MALE JOURNALIST (EAST AFRICA)

Perception of platforms' practices and responses to requests

As reported above, most HRDs and journalists who reportedly experienced online threats or attacks did not contact social media and communication platforms to report or seek information about them. More tried to reach platforms when their accounts or profiles were blocked, or when specific content was removed. Levels of satisfaction with platform response varied, and most remain unaware of the reasons for reported restrictions. The following key concerns were highlighted in interviews:

A) **Cumbersome procedures -** Contacting platforms to request the restoration of content or accounts can be a time-consuming and complex process with uncertain results. Interviewees underlined that the processes can be drawn out for a couple of months, often without yielding any tangible results. The perceived challenges in reporting content and account blocking or removal through platform channels seem to drive several HRDs and journalists to seek support from protection networks or other specialized organizations that have established some form of contact with platforms and can therefore assist in presenting requests or obtaining quicker responses.

> *"Restoring blocked accounts is really demanding and time consuming, and it can take months. I have tried to contact the platforms myself to have blocked accounts restored and did not have a good response. In our organization we always have multiple social media accounts in case some get blocked. [...] Usually when activists with many followers and high visibility get their accounts blocked, they ask for help through specialized NGOs and protection networks."*
>
> TESTIMONY - NGO WORKERS AND HRDS (MENA)

> *"[...] Sometimes social media block your posts and only inform you that the publication violates platform rules, without any further detail. There are different rules in different platforms. We as activists in our community usually post the same content in different platforms, on some it gets blocked, on others it is published. [...] I don't have the time to engage in conversations with the platforms each time my posts don't get through. It is really time consuming and frustrating."*
>
> TESTIMONY - MALE HRD (EAST AFRICA)

B) **Lack of accessible or efficient interlocutors –** Several respondents raised concerns about the lack of simple and safe channels to contact social media platforms in the locations where HRDs and journalists are based. Some interviewees emphasized the impact of what they considered to be very limited resources allocated to work in their countries. Repeated references were made in interviews to how platforms were ill-equipped to capture context-related nuances and provide adequate responses to concerns about hate speech or disinformation in the context of ongoing armed conflict and regular violent threats.

C) **Language -** Around a quarter of the surveyed participants responded that language barriers affected their engagement with platforms. Interviewed groups noted the overreliance of moderation on automated systems, expressing repeated concerns about their inefficiency when dealing with non-English languages.

In East Africa, participants made specific reference to the very harsh working conditions imposed on workers who provide human revision of automated moderation in the region as an example of the lack of investment and attention to the regional context.

Participants also indicated how the enforcement of content moderation sometimes resulted in self-censorship or language adaptation to avoid automated systems barriers. As a result of experiencing repeated restrictions, some respondents indicated that they avoid using certain terms (such as

> *"When I started working for the NGO, I was very active on social media, posting everything. However, after 8 months, I began to receive threatening phone calls from the intelligence office, which led me to reduce my social media activity. Now, when I post on social media, I'm aware that I'm under constant surveillance, so I maintain a low profile and only publish content through the NGO's accounts. To avoid having our posts rejected or our accounts blocked by the platform, we need to be careful about what we post, the language we use, and the types of pictures we share. We often face moderation issues on social media when our posts include the word "shaheed" (i.e., martyr) or pictures depicting human rights abuses. These moderation policies force us to self-censor so that our posts can be published on social media."*
>
> TESTIMONY - FEMALE NGO WORKER AND HRD (MENA)

"shaheed," an Arabic word which translates as "martyr"), refrain from displaying certain images, or choose not to publish politically sensitive content to prevent rejection by platforms.

D) **Limited transparency and bias –** Several interviewees underlined that they encountered limited transparency and inconsistencies in content moderation, including a lack of responsiveness from social media platforms regarding their interventions. According to the same interviews, rulings and appeals processes showed inconsistencies, and platforms often failed to provide justifications or offered generic or insufficient explanations.

Interviewed groups also perceived bias favouring demands for content blocking or removal coming from Government authorities or other powerful groups. Some respondents illustrated concerns about the perceived bias of content moderation against Palestinians in the context of hostilities between Palestine and Israel.

> *"Relationships between social media platforms and the government are not public, and this creates a lot of uncertainty and distrust. As journalists, we try to work with these platforms, but it seems like they don't invest much in maintaining proper moderation teams in our country. Instead, they appear more focused on building relationships with the government. Our mistrust in the government starts to become mistrust in the companies themselves, even though we still rely on them and hope they'll be "different". This lack of trust in social media companies affects us as journalists, making us more cautious about how we express ourselves and leading us to self-censor our publications in social media."*
>
> TESTIMONY - GROUP OF JOURNALISTS (EAST AFRICA)

E) **Limited understanding of critical contexts -** HRDs and journalists working in countries experiencing armed conflict and severe political turmoil reported concerns about the impact of stringent moderation and restrictions on the information, data, and images they are allowed to publish. Measures blocking the publication of photographic material or footage related to conflict or war due to violations of platform rules have reportedly affected their capacity to document and disseminate information on human rights violations. The limited capacity of platforms to respond to demands by civil society actors operating in volatile contexts was also raised as a significant risk factor that exacerbates protection concerns for HRDs and journalists due to delayed or inefficient responses to reported threats.

> *"In our organization, we work with cases of torture and forced disappearance in the context of conflict. Social media platforms are extremely useful for connecting with victims' families, conducting outreach, and documenting cases with pictures.*
>
> *Using pictures on social media can be very complicated due to their nature, as they often don't pass moderation standards because they're usually associated with conflict and violence. However, these images, even if they depict violence, abuse, and sad stories in general, are essential to our work. They help us identify victims and their families, as well as document violations of humanitarian law and human rights."*
>
> TESTIMONY - FEMALE NGO WORKER AND HRD (MENA)

F) **Limited efficiency of protection by platforms** - Participants indicated that, while some platforms have engaged directly with protection networks and actors and grant privileged access to speedy treatment of reported cases of incitement and other significant concerns, the large majority of HRDs and journalists interviewed were not aware of any platform initiatives or tools. Among those who were familiar with some type of assistance channels, many noted that the tools were not easily accessible, especially due to the location of platform staff (in a different country, possibly another region) or lack of network affiliation. Others criticized the lack of more proactive responses to severe risks and what they perceived as platforms outsourcing responsibilities to civil society to provide early warnings about critical incitement situations, which platforms would otherwise fail to understand or recognize.

Overview of civil society asks of platforms

Besides collecting information on HRDs' and journalists' overall experiences with social media and communication platforms, the pilot study invited all surveyed and interviewed participants to list what they identified as important steps platforms could take to improve their experiences online and to deal with the reported concerns. The demands listed by participants mostly reflect the concerns identified by the survey questions but also reflect ongoing debates about overall challenges faced by civil society actors while using social media and communication platforms, particularly in countries of the Global South.

The level of detail of demands presented mostly corresponded to the interviewees' degree of exposure to debates on the quality of moderation. Participants with limited exposure to platforms' practices frequently emphasized the need for capacity building and direct support, while those already familiar with content moderation mechanisms provided more detailed recommendations about assistance processes and responses to users' requests. Some of the most recurrent demands are grouped below.

A. Enhance cooperation with organizations based in country

Platforms should establish or enhance engagement and cooperation with HRDs, journalists, and protection networks based within MENA and East Africa, paying special attention to those operating in high-risk contexts. Companies should cooperate with local civil society actors in order to (*inter alia*):

(a) Design and review the implementation of platforms' policies and standards, including content moderation standards, paying particular attention on their impact on local HRDs' and journalists' expression;
(b) Identify relevant protection gaps with a view to adopting special initiatives or adapting policies to specific contexts or critical periods (such as electoral contexts, contexts of elevated violent risk or recurrent risks of detention and judicial harassment of civil society actors);
(c) Undertake reviews of the impact of online threats and attacks against HRDs and journalists, including special attention to gendered elements and repercussions for minorities; and
(d) Promote outreach and awareness raising on platform policies, decisions and programmes among HRDs and journalists.

B. Specific asks for direct support

Interviewees emphasized the relevance of platforms establishing special features or privileges to address HRDs' and journalists' demands more quickly and efficiently, taking into consideration that they face greater risks than most users. Suggestions for potential support included:
(a) Establishing or improving systems of account verification;
(b) Providing enhanced privacy and security guidance and options, including specific notifications about potential hacking attempts, targeting, and surveillance by state and non-state actors;

(c) Establishing or improving channels that offer priority access to moderation and appeal mechanisms; and

(d) Offering reduced fees or exemptions when using commercial tools to promote human rights advocacy.

Interviewees also emphasized the need for companies to provide direct guidance and support for HRDs and journalists using their platforms, including through training materials and opportunities for groups operating in different contexts to improve their risk awareness and cyber protection skills.

C. Improved transparency and redress mechanisms and resistance against arbitrary asks

Given their perceptions of platforms' bias and sometimes arbitrary or inefficient content moderation practices, respondents emphasized the need for greater clarity about companies' practices and the ways they relate with authorities and respond to State requests. References were also made to the need for increased resistance against pressure and the need to expand investment in the implementation of platforms' policies in different locations and languages to meet minimal quality standards. Participants' responses included demands for platforms to:

(a) Expand transparency at all stages of operations, systematically providing data on the implementation of policies and practices related to sensitive areas for HRDs' and journalists' work and safety (such as data protection and content moderation), and systematically enabling the review of their practices;

(b) Expand human and linguistic capacity to provide civil society actors with more accessible channels for reporting and to review the implementation of content moderation and protection measures;

(c) Increase investment in initiatives that track and document coordinated attacks, including reporting on their frequency and consistency, to allow for a better understanding of major threats affecting civil society, particularly during critical periods;

(d) Take all lawful measures to resist State demands for restricting communications and collecting private data of HRDs and journalists, and strive to enable full and immediate disclosure of information about all requests from authorities and, in particular, inform civil society actors directly affected in more efficient ways; and

(e) Provide effective remedy for HRDs and journalists who experience restrictions due to technical errors, wrongful, or unfair moderation decisions.

---

[i] For OHCHR, Middle East and North Africa includes 19 countries: Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Occupied Palestinian Territory, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates and Yemen. East Africa includes 8 countries: Djibouti, Ethiopia, Kenya, Tanzania, Somalia, South Sudan, Sudan and Uganda.

[ii] *An online survey with anonymous and mostly -closed-ended questions (with space for additional text to be shared) was implemented between December and February. The survey was anonymous and shared directly with organizations providing protection to journalists and HRDs in both regions with versions in English and Arabic.*

[iii] Attackers sending fraudulent messages designed to trick HRDs or journalists into revealing passwords or other sensitive information.

[iv] Hostile actors presenting themselves as the HRD or journalist to manipulate debate or obtain information.

[v] Publication of private information with malicious intent.