European Digital Media Observatory

SCHOOL OF TRANSNATIONAL GOVERNANCE

European Digital Media Observatory

EDMO Survey Mapping: adversarial behaviour, prevention and protection

December 2024

EDMO Survey Mapping: adversarial behaviour, prevention and protection

EDMO Survey Mapping: adversarial behaviour, prevention and protection

Introduction

The EDMO Survey on Adversarial Behaviour, Prevention and Protection was conducted in response to an increasingly uninhibited climate of online and offline abuse in Europe, as in many other regions around the world, directed at academic researchers, fact-checkers, journalists and other professionals working to better understand and counter online disinformation.

As a multidisciplinary network that brings together stakeholders from the research, fact-checking and media literacy communities with specific expertise in the area of disinformation, EDMO and its fourteen regional and national Hubs (the EDMO Network) are well-placed to contribute to the mapping of online and offline abuses against professionals in the field, with the primary goal of raising awareness of the phenomenon and of the responses in place, or lack thereof.

The EDMO Survey, conducted from 25 July to 2 October 2024, provides an overview of adversarial behaviour experienced by individual members of the EDMO network, as well as of the prevention and protection mechanisms in place in the respondents' organisations.

Scope and methodology

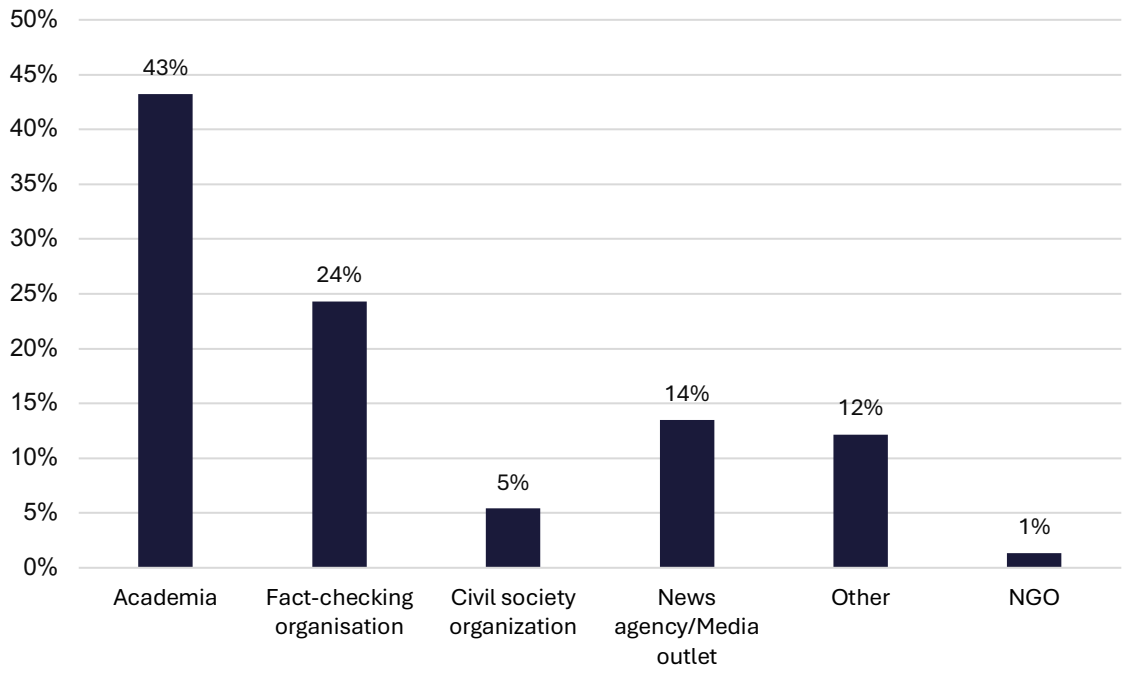
The survey questionnaire, comprising 19 questions was structured into four sections i) About you with 5 questions concerning the respondents' professional role, type of institution and country of work; ii) Adversarial Behaviour with 9 questions on the modality, type, triggering factors and responsible entities for the reported abuses as well as the frequency and eventual involvement of local authorities; iii) Preventive measures with 3 questions regarding the potential future attacks, preventive measures in place and recommended ones as well as iv) Protective measures with 2 questions on those in place and recommended ones. Questions allowing for additional comments were also foreseen. EDMO sent the survey on 25 July to all regional and national EDMO Hubs and to all members of the EDMO governance bodies.

Highlights

With 74 responses, the survey captures insights from a diverse range of roles and institutions across the network including academia, fact-checking organizations, civil society organizations, and media outlets.
It is worth highlighting that a substantial majority of respondents (66%) reported experiencing adversarial behaviour. The five most frequent forms of adversarial conduct are online harassment (Trolling), threatening and intimidating emails, coordinated harassment (Brigading), physical threats or harm and doxing. The primary triggers for the reported abuses include retaliation against published work on contentious topics, general distrust and hostility towards the community at work to counter disinformation, and public speaking, especially about conspiracy theories or digital platform regulation. Adversarial actions are often perpetrated by individuals, ideologically motivated groups, alternative media outlets, and political or government-related entities. While sporadic occurrences are most common, several respondents report frequent abuses, particularly following the publication of work on controversial issues.

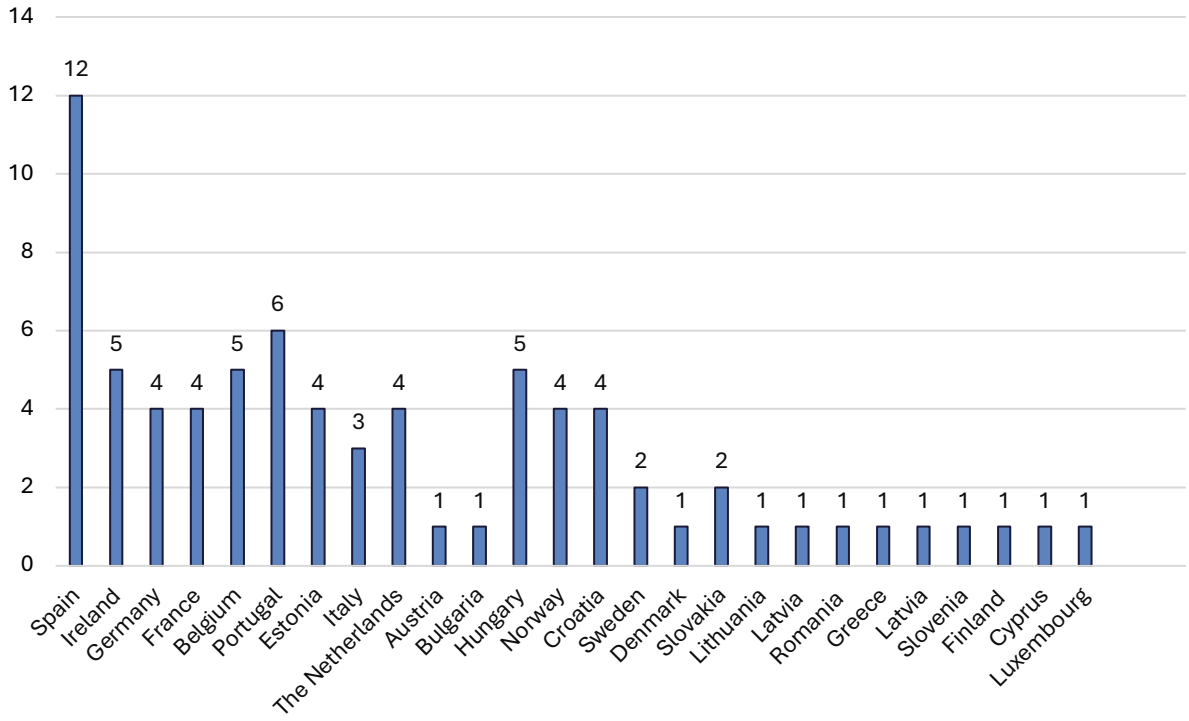
Section 1: About the respondents

Role and institution

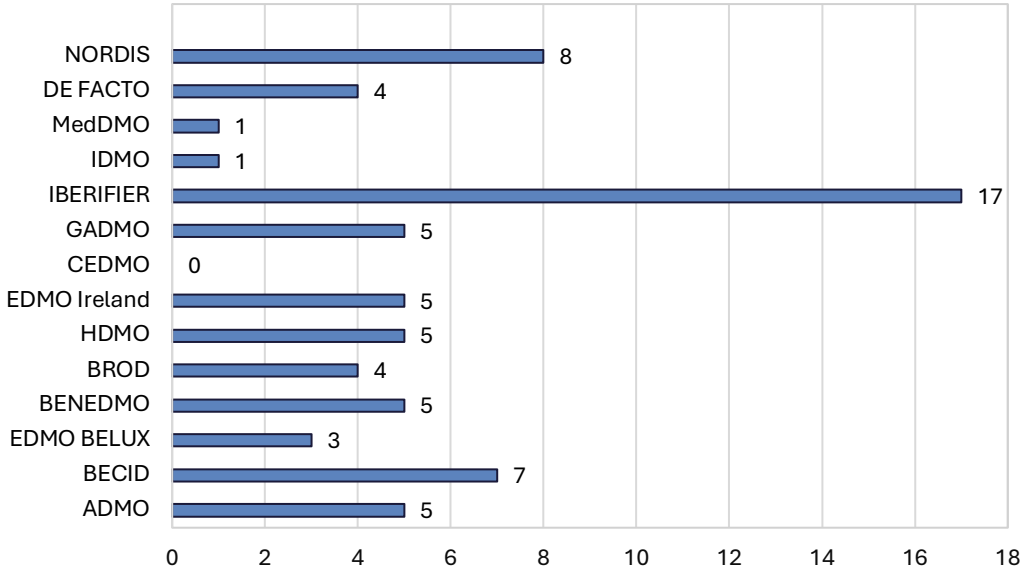


Regarding the affiliation of the 74 respondents, 43% work for academic institutions, 24% work for a fact-checking organisation, and 14% for a news agency or media outlet. Among other affiliations, technology companies and media literacy organisations were mentioned.
In terms of roles, most respondents hold academic positions, often linked to coordination roles in the EDMO hubs and/or leadership positions in research projects. A significant number of respondents are journalists and fact-checkers, followed by media literacy experts and technology consultants.

Country of work



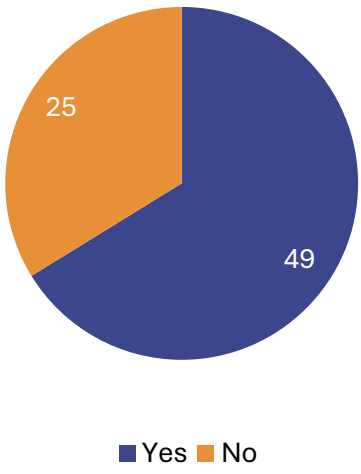
The average number of respondents per country is 3. There is a high representation of respondents that operate in Spain (12) and Portugal (6) as well as a relevant number of countries that count 4 to 5 respondents respectively Ireland, Belgium, and Hungary (5) and Germany, France, Estonia, the Netherlands, Norway and Croatia (4). A higher rate of respondents per country can indicate stronger engagement with the issues addressed by the survey but cannot be taken as an indicator of the overall distribution of adversarial behaviour across the network.
Responses from only 1-2 persons per country can reflect fewer reported incidents or the fact that a smaller number of institutions from these regions are involved in the EDMO network.



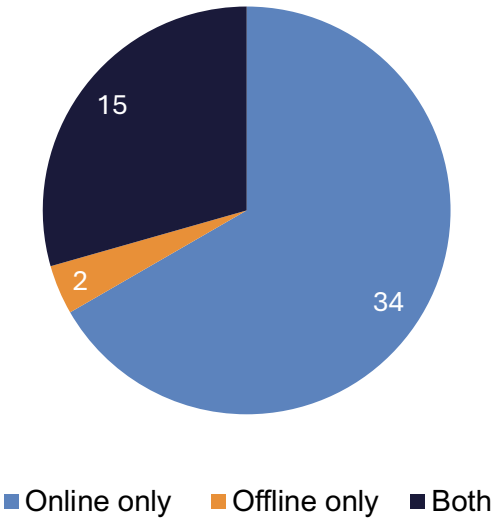
Section 2: Adversarial Behaviour

Overall, a total of 160 instances of adversarial behaviour were reported with a great variety. This underscores the widespread nature of the phenomenon within the EDMO network.

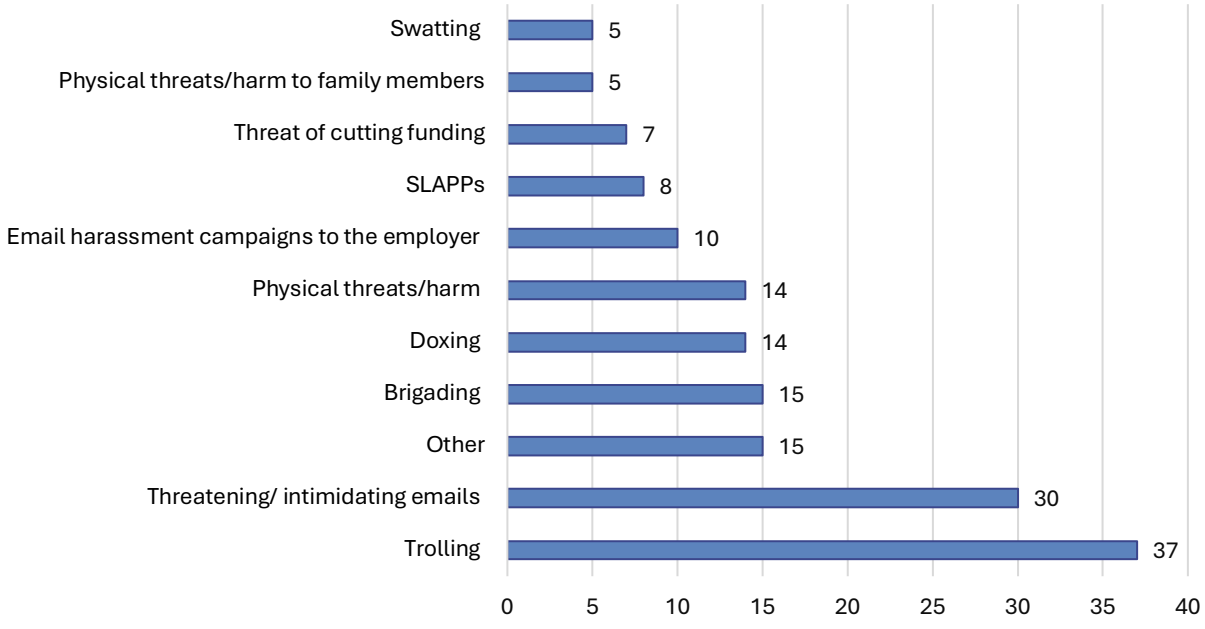
A strong majority, 49 out of 74 respondents (approximately 66%), reported having been the target of adversarial behaviour including attacks or threats. While 25 replied negatively to the question from their responses it emerges the extent to which they consider that adversarial behaviour is a source of concern for the whole EDMO community.



Most adversarial behaviour is experienced in the online environment only (34) while a significant amount reported being attacked both online and offline (15). Offline-only incidents are very rare (2).



Type of adversarial behaviour experienced

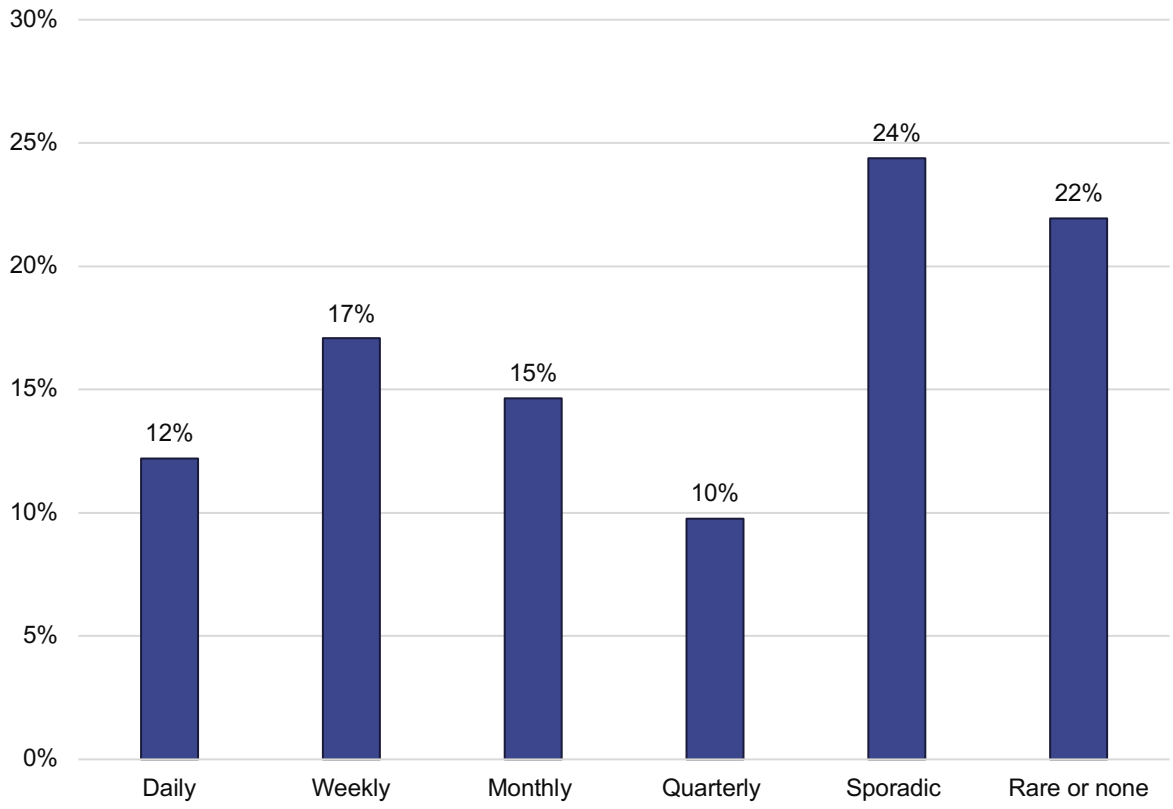


Trolling (harassing someone online with the intent of provoking a negative reaction); Brigading (when a group of people work together to harass an individual); Doxing (revealing personal information publicly without consent); Swatting (initiating a law enforcement response on an unsuspecting victim); SLAPPs (strategic lawsuits against public participation).

Among a given set of adversarial behaviour, the most frequently reported form of abuse with 37 instances is online harassment aimed at provoking negative reactions (**Trolling**). The second most common type of attack, with 30 cases, is threats and intimidation via email (**Threatening/Intimidating Emails**). Other trends include 15 cases of coordinated harassment by groups (**Brigading**), 14 cases, plus 5 additional instances involving family members, of **physical threats/harm**, 14 instances of personal information being revealed without consent (**Doxing**), 10 cases of **email harassment campaigns to employers** with potential professional repercussions, 8 cases of strategic lawsuits against public participation (**SLAPPs**), 7 reported instances of financial intimidation (**Threat of cutting funding)** and 5 which involve initiations of law enforcement responses on unsuspected victims (**Swatting**).

In addition to these main categories, respondents have also highlighted other threats including **discrediting research** through accusations of partisanship and not respecting the principle of neutrality; **cyber-attacks** such as DoS (Denial of Service) attacks, phishing, and server attacks; **harassment by political groups or state actors** entailing false and misleading media coverage and framing organizations as “foreign agents” as well as intimidation through political inquiries and wiretapping.

Frequency of adversarial behaviour in the past year



When asked about the frequency of adversarial behaviour in the past year, the most common trend was **sporadic occurrences** (2-3 times/year), reported by 25% of respondents. Additionally, 22% experienced **rare or no** adversarial actions. **Weekly** (200-50 times/year) and **monthly** (50-12 times/year) occurrences were noted by 17% and 15% of respondents, respectively, indicating a moderate level of regularity. **Daily** occurrences were relatively low at 12%, suggesting that such behaviour is not predominantly a daily issue. While most respondents experienced isolated incidents, a notable number reported higher frequency adversarial behaviour, often in waves following the publication of work on sensitive topics.

Triggering factor (if aware)

The main trends among the triggering factors mentioned by the respondents include:

- **Retaliation against published work**, especially fact-checking touching upon contentious topics such as disinformation related to public health and Covid-19, migration, the environment and women. Investigative work into far-right groups and/or online conspiracy theories networks was also mentioned as a common trigger.

- **General hostility for the community working to counter disinformation,** especially if mentioned by national authorities within their strategy to address the phenomenon. A negative perception of the work of EDMO and the Hubs labelled as endangering freedom of expression or as serving foreign interests is also mentioned, especially when EDMO's work is erroneously understood as content moderation or framed as censorship.

- **Public presentations** at events are also mentioned as triggering adversarial behaviour, particularly when the work presented challenges conspiracy theories or addresses the regulation of digital platforms including for empowering researchers with data access.

The above-triggering factors build upon recurrent allegations as also highlighted by the Observatory of Disinformation Narratives Against the Media. This includes labelling members of the community working to counter disinformation as:
"Censors" or "Ministry of Truth" not committed to facts or transparency but enforcers of an authoritarian agenda that suppresses free speech and controls the flow of information. This narrative is directed especially at fact-checkers to delegitimize their professional reputation and present them as active agents of conspiracy against people.
"Foreign Agent" serving foreign interests – sometimes together with an "establishment" beholden to foreign powers – rather than those of their own country. This discourse can be especially dangerous if it justifies attacks against these experts as "enemies" of the

national society threatening the “freedom of speech” on behalf of or with the support of outside forces.

Recipient of **“Dubious finances”.** Questioning the legitimacy and transparency of an organization’s funding often suggests that the work of such an organisation is inherently compromised due to funding sources i.e., foreign governments, powerful global elites, or other organizations with specific political agendas. Accusations therefore imply that the work produced by such organisations serves hidden or vested interests rather than the public good.

Politically biased, aligning with specific political ideologies rather than adhering to professional ethical standards. The work of the community can be portrayed as ideologically driven rather than fact-based, supporting political activism either with a “left-wing” agenda, especially in Western Europe, where far-right and conspiracy groups have grown in recent years, or a “liberal agenda”, often in Central and Eastern Europe, where Russian or Russian-affiliated sources are the main purveyors of disinformation.

Responsible entity (if aware)

Among the entities mentioned by respondents as responsible for the adversarial behaviour the most representative trends include:

- **Individuals** either random users or prominent figures in the political and or business environment as well as social media influencers.

- **Ideologically motivated groups** (e.g., far-right) **or conspiracy theory communities** purveying disinformation, especially around COVID-19 and Russia.

- **Alternative/fringe media outlets** including specific mentions of far-right blogs, pro-Kremlin outlets and local television channels.

- **Political and government-related entities** especially if being the object of respondents’ investigative work. State actors and government-related mediums were particularly mentioned in Hungary as well as in Slovakia as responsible for harassment and other adversarial actions.

In some cases, respondents were unable to identify the specific entities responsible for adversarial behaviour.

Local law enforcement authorities

Regarding the involvement of local law enforcement authorities in addressing adversarial behaviour, the main trends reported are:

- **No involvement**

- **Police involvement in specific but severe cases** such as reporting death threats, patrolling places of residence and protecting respondents when conducting awareness-raising campaigns on the ground.

- **Legal actions and complaints** filed with the courts following very serious attacks including death threats or intimidations against family members. Some respondents reported that their complaints did not lead to charges or decisions against the offenders.

- **Informal consultations** signaling threatening emails or social media posts to authorities without making formal complaints.

Other comments

Among the additional comments, respondents mentioned that most adversarial conducts are directed at the organization rather than the individuals and that isolated incidents are more common than coordinated and regular attacks.

Some organizations shared that higher levels of harassment were reached during the Covid-19 pandemic and others reported that attacks worsened after entering into collaborations with platforms for fact-checking services.

In some member states such as Ireland, constructive developments were reported including the setup of dedicated hotlines between media professionals and police authorities with quarterly meetings to discuss threats and responses.

In other member states, opposing trends were highlighted such as raising the level of surveillance and endangering freedom of expression including media freedom. The adoption in early 2024 of the '**Sovereignty Protection Act**' in Hungary is a fitting example, with the establishment of a Sovereignty Protection Office (SPO) acting to discredit independent media, accusing it of spreading disinformation in service of foreign interests.

Some respondents underlined how they could turn their experiences of online threats and harassment into awareness-raising material to strengthen the resilience of other colleagues in the face of potential future attacks.

Section 3: Preventive measures and protocols

Potential future adversarial actions

When asked about potential adversarial actions that respondents could envisage experiencing in the future, many expressed uncertainty about specific future abuses but emphasized the need to be prepared for various scenarios. In particular, establishing networks for support and information sharing for best response strategies was suggested.

The main trends for the type of potential future adversarial behaviour reported were:

- **Continuation of current adversarial** behaviour including harassment from political representatives and state actors.
- **Potential escalation from online to offline** physical threats or attacks including during demonstrations, intrusions to headquarters and spyware used against them by state authorities.
- **Increased legal adversarial actions** (e.g., lawsuits for defamation) and investigations against journalists.
- **Resources challenges** with potential cuts in funding driven by influential people or political pressure.
- Concerns about deepfakes and **AI-generated false images** were also mentioned.

Preventive measures and protocols

When asked about the preventive measures and protocols currently in place in the respondents' organisations, a relevant number reported the absence of such measures while underlying the relevance of mapping the gaps to find collective solutions.

Preventive measures currently in place fall within the main trends highlighted below:

- **Security** and digital safety including implementation of new security systems at office premises and digital security measures such as spam filters, migration to more secure networks and obtaining security certifications.

- **Guidelines** for handling harassment involving topics such as mental health and well-being, online and offline security, including for public incidents.

- **Monitoring** incidents to track frequency, identify patterns and assess the level of threats to address any suspicious or harmful behaviour.

- **Anonymity** for staff including not signing articles with individual names and applying protocols for non-disclosure of protected personal information.

- **Neutral communication** and closing comments on controversial topics.

A number of measures are only implemented in some member states and/or partially. Therefore respondents recommend the need to further develop and streamline the implementation of specific actions including:

- **Security measures** at office premises and **cybersecurity measures** such as using only organization-provided laptops and devices. Implementing watermarks for AI-generated content was also mentioned.

- **More developed guidelines** on best practices to help reduce the intensity and negative impacts of harassment should it arise.

- **Enhanced monitoring and evaluation** of potential harassment and threats

When comparing respondents' answers for preventive measures already in place in their organisations and recommended measures to put in place, two elements that stand out as desirable preventive measures are:

- **Peer-to-peer experience sharing** within and among organizations to learn best practices and establish support networks.

- **Raising awareness** among the public and policymakers about adversarial behaviour's prevalence and impact on the community.

Section 4: Protective measures and protocols

Protective measures and protocols

While some mentioned the lack of established protective measures in their organisations, the main measures reported fall within the below categories:

- **Archiving evidence** of adversarial behaviour.
- **Legal support** for filing complaints with relevant authorities when necessary.
- **Psychological** support and mental health hotlines for employees.
- **Training** both general and specific for newsrooms and fact-checkers on how to respond to threats and harassment including in terms of communication.
- **Crisis response protocol** for adversarial situations.
- **Peer-to-peer support systems** including the appointment of internal referees or task forces ready to support in response to harassment situations.

When asked about recommended protective measures, the main suggestions highlight the need for deepening actions such as:

- **Stronger and more rigorous enforcement** of existing legal frameworks**.**
- **More robust legal advice** to employees for handling defamation, insults, and slander.
- **Enhancing support systems** for targeted members of the community and exchange of best practices with experienced organizations in the field.

Conclusion

The survey identified the prevalence of adversarial actions against members of the EDMO network. More than half of the respondents (66%) reported experiencing abuses, of which online harassment was the most common type. These results underscore the persistent and growing risks that the counter-disinformation community faces, both in Europe and beyond.

Among the recommendations for preventing such risks, respondents highlight strengthening networks for peer-to-peer support and information sharing; thorough monitoring and evaluation of threats; developing more robust guidelines on how to prepare in the event of adversarial actions; enhancing security both online and offline and strengthening activities to raise public and policymaker awareness of the issue.

Recommendations for enhancing protection from abuses include stronger and more rigorous enforcement of existing legal frameworks, more robust legal advice and support for employees, and strengthened peer-to-peer support systems for targeted members of the community.

EDMO calls on all competent authorities to address this alarming challenge with the greatest determination, in line with the Charter of Fundamental Rights of the EU. EDMO and its network of national and regional Hubs will build on the insights and recommendations highlighted in this survey. Further consultations with the network will be carried out to identify both preventive and protective measures that could contribute to mitigating the risks faced by its community of practitioners.

www.edmo.eu

The European Digital Media Observatory has received funding from the European Union under contract number LC-01935415