ONLINE SAFETY AND DIGITAL SECURITY FOR *ALL* JOURNALISTS:

A PREREQUISITE FOR MEDIA FREEDOM

Author: **Arzu Geybullayeva**

Report Coordinator: Julia Haas and Cecilia Lagomarsino

This report was commissioned by the OSCE Representative on Freedom of the Media as part of the project to put a spotlight on AI and freedom of expression, #SAIFE. The views, findings, interpretations, recommendations and conclusions expressed herein are those of the author and do not necessarily represent the official position of the OSCE and/or its participating States.

Wallnerstrasse 6
A-1010 Vienna, Austria
Tel.: +43-1 514 36 68 00
E-mail: pm-fom@osce.org
http://www.osce.org/fom/sofjo

TABLE OF CONTENTS

GLOSSARY OF TERMS

Digital security

Umbrella term used to describe protection mechanisms and practices for online identity, data, information networks and devices-in-use (such as emails, computers, tablets, phones, social media accounts, medical and bank records, communication apps, and so on).

Online safety

Online safety can be understood as the absence of *online harassment and abuse*, a phenomenon with many names: cyber harassment, cyberbullying, trolling, flaming, etc. Generally, it can be defined as the "pervasive or severe targeting of an individual or group online through harmful behavior."[1]

Types of online harassment can be manifold and multifaceted. They include, inter alia, astroturfing[2], Denial of Service Attacks (DDoS), sharing of nonconsensual intimate images, online impersonation, online sexual harassment, targeted disinformation, doxing[3], phishing, cyberstalking, deepfakes, hacking.[4]

As digital security and online safety are closely related, partly overlapping and mutually reinforcing concepts, it is difficult to make a clear distinction between both terms. The above definitions attempt to illustrate what the terms typically refer to, and how the terms are used for the purpose of this paper.

EXECUTIVE SUMMARY

The internet has brought extensive opportunities for people worldwide; however, it has also brought challenges to its users on a global scale. From governments relying on surveillance tools and the internet to criminalize critical voices, to companies putting profits before human rights, while failing to offer meaningful protection practices to its users from harassment, hate and violence they face on the platforms operated by these companies.[5] The latter, specifically online threats and harassment faced by journalists has become "the new frontline for journalists' safety".[6] Especially affected are women journalists and media practitioners from other marginalized groups.[7]

[1] PEN America Online Harassment Manual, Defining "online abuse": a glossary of terms, https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms.
[2] A practice of masking the sponsor of a message to make it appear stemming from grassroots participants, or mimicking organic reactions.
[3] Sharing private or personally identifiable information.
[4] PEN American Online Harassment Manual, https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms.
[5] Katie Porter, David Kaye, The UDHR, Digital Authoritarianism, and Human Rights after Trump, https://www.justsecurity.org/73785/the-udhr-digital-authoritarianism-and-human-rights-after-trump.
[6] UNESCO, ICFJ, Online violence against women journalists: a global snapshot of incidence and impacts, https://unesdoc.unesco.org/ark:/48223/pf0000375136.
[7] IWMF, Attacks and Harassment: the Impact on Female Journalists and Their Work, https://www.iwmf.org/attacks-and-harassment.

According to a 2020 ICFJ-UNESCO global study, nearly three in four women journalists experience online violence. 30 per cent of the respondents indicated they self-censor following online violence, and almost 40 per cent retreat from visibility.[8] An International Women Media Foundation study showed that a third of women journalists who have experienced some form of threat, attack or harassment online, "have considered leaving the profession due to online attacks".[9] But the trauma and impact on the individual level is not the only consequence. According to Women's Media Center, "in addition to clearly influencing how journalists work, online harassment also affects organizations' ability to recruit, retain, and reward diverse staff and cultivate inclusive media environments and leadership".[10]

An early and still often promoted strategy to cope with growing trends in online harassment has been to invest in digital security training and awareness raising among journalists and media practitioners at large. However, as this paper shows, the shifting of focus on digital security protocols only, cannot serve as a sustainable and long-term solution model for combating online threats and harassment. Instead, existing approaches, practices as well as gaps should continuously be re-evaluated to mitigate the risks to online safety. Demands for greater transparency and accountability on behalf of companies and platforms are equally important, as are long-term commitments to eradicate online harms by national governments, international human rights organizations, freedom of the media advocates, newsrooms and others. Otherwise, harms done online will continue to hamper plurality, media freedom and hinder the safety of journalists.

INTRODUCTION

The good news: awareness among journalists worldwide on digital security as a concept and its use to counter digital threats has improved. Much of this awareness on privacy, security and surveillance is due to Edward Snowden revelations in 2013 which played an important role in shifting perceptions among media practitioners and the societies more broadly, of privacy, security, and surveillance.[11] Since then, while digital threats continue to affect the lives of journalists across the globe, there is certainly a far better understanding of the types of digital threats and targeting methods deployed by states or malicious actors. This overall awareness of types of threats also helped develop various resources (such as digital security toolbox and manuals) and recommendations by digital security experts for media practitioners and for civil society more broadly.

[8] International Center for Journalists and UNESCO Global Study: Online Violence Against Women Journalists, 2020, https://www.icfj.org/our-work/icfj-unesco-global-study-online-violence-against-women-journalists.
[9] IWMF, Attacks and Harassment: the Impact on Female Journalists and Their Work, https://www.iwmf.org/attacks-and-harassment.
[10] Women's Media Center, What online harassment tells us about our newsrooms: from individuals to institutions, Online News Association 2019 Conference, https://womensmediacenter.com/assets/site/reports/what-online-harassment-tells-us-about-our-newsrooms-from-individuals-to-institutions-a-womens-media-center-report/WMC-Report-What-Online-Harassment_Tells_Us_About_Our_Newsrooms.pdf.
[11] Camille Fassett, "How Snowden has changed journalism and privacy, five years later," Freedom of the Press Foundation, June 6, 2018, https://freedom.press/news/how-snowden-has-changed-journalism-and-privacy-five-years-later.

As a result, today, there exists, a global effort acknowledging the ever so pressing digital threats, and working towards minimizing these threats. At the same time, journalists and newsrooms continue to document the extent of digital authoritarianism tools deploying information controls worldwide.[12]

The bad news is that the threats[13] have become a norm, they have become consistent, overt and more journalists are subject to malicious behavior[14] online on a regular basis. Numerous research and documentation[15] attests that the proliferation of such threats targeting journalists have reached an unprecedented level, with an overall trend of growing digital authoritarianism practices, often combined with non-digital measures[16] that make the work of media practitioners challenging, dangerous and sometimes impossible all together.

In addition to what by now has become the usual methods[17] of online targeting, such as online harassment, disinformation, account hijacking, DDoS attacks, hacking and phishing attempts, certain actors, including governments, are resorting to new measures. These novel measures include but are not limited to changing laws and regulations, thus, shifting restrictions to online spaces, and contributing to an environment of fear. As a result, the basic understanding of digital security tools and threat modelling that are often based on less pervasive attacks are lagging behind, making it much harder for those targeted to mitigate risks, while continuing their work in a highly hostile environment.

---

[12] Digital tools of information control include internet shutdowns, government imposed restrictions on online communication internet throttling, deployment of surveillance technology to disrupt flow of independent information, by blocking access to independent websites, social media platforms and communication apps. "Digital technologies as a means of pressions and social control," Dorota Glowacka, Richard Youngs, Adela Pintea, Eweline Wolosik, European Parliament, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653636/EXPO_STU(2021)653636_EN.pdf, April 23, 2021. See also "The use of digital information technology by authoritarian regimes to surveil, repress, and manipulate domestic and foreign populations," Alina Polyakova and Chris Meserole, Policy Brief, "Exporting digital authoritarianism", Brookings, https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190826_digital_authoritarianism_polyakova_meserole.pdf.

[13] "(i) internet shutdowns and other network disruptions, as well as mass and targeted surveillance; (ii) an increasing use of the 'next generation repression toolkit', which encompasses practices that are more difficult to detect and hold accountable for (e.g. government hacking or state-sponsored online harassment campaigns); (iii) the expansion of digital authoritarian practices outside national borders through targeting diaspora or the export of surveillance technology. The rising power of a handful of tech companies which have become the gatekeepers of fundamental rights in the digital realm poses yet another significant challenge to those rights." From "Digital technologies as a means of pressions and social control," Dorota Glowacka, Richard Youngs, Adela Pintea, Eweline Wolosik, European Parliament, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653636/EXPO_STU(2021)653636_EN.pdf, April 23, 2021.

[14] Malicious behavior in this context implies to trolling, hacking, account take down requests, fake copyright violation reports to social media platforms, impersonations, doxing, DDoS attacks, etc.

[15] See, for example, "Digital technologies as a means of pressions and social control," Dorota Glowacka, Richard Youngs, Adela Pintea, Eweline Wolosik, European Parliament, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653636/EXPO_STU(2021)653636_EN.pdf, April 23, 2021; Erol Yayboke, "Promote and Build: A Strategic Approach to Digital Authoritarianism," Center for Strategic and International Studies, October 15, 2020, https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism; and Alina Polyakova, Chris Meserole, "Exporting digital authoritarianism," Foreign Policy at Brookings, August 2019, https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190826_digital_authoritarianism_polyakova_meserole.pdf.

[16] Offline measures including traditional forms of intimidation, arrests, detentions, physical violence.

[17] Online harassment field manual, PEN America, https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms.

The aim of this document is to expand on these realities. This report argues that while digital security measures for journalists cannot eradicate the harassment and abuse online alone, taking stock of digital security concepts, manuals, and measures, and adopting a holistic and intersectional approach to digital safety can minimize the damage. This means that tools and recommendations need to be holistic and combined with other mechanisms, such as holding the perpetrators of these threats to account, mitigating the risks of being targeted online, and preventing attacks from happening again.

In that regard, this document acknowledges the gap that in order to sustainably and effectively address abuse online and other forms of digital threats, the responsibility not only falls on media practitioners, but on all stakeholders involved, including journalists, newsrooms, companies, platforms, governments, and international institutions. Acting together and closing this gap requires international mechanisms to be in place for actors that do not comply with global efforts, measures and recommendations, including within the OSCE and among its participating States.

A HOLISTIC APPROACH IS NEEDED

The growing influence of artificial intelligence (AI) technologies entail the possibility of far more pervasive targeting and attacks online and in the *metaverse*,[18] as recent experience shows , especially affected are women journalists and media practitioners from other marginalized groups.[19]

The latter is no new revelation. Ever since the Gamergate scandal in 2014,[20] the targeting of women and marginalized groups online has been widely discussed, documented and reported on. According to the SOFJO (Safety of Female Journalists Online) Resource Guide released by the OSCE Representative on Freedom of the Media (RFoM) in 2020, the targeting of women journalists and marginalized voices online has manifested itself in "direct or indirect threats of physical or sexual violence, offensive messages, and targeted harassment (often in the form of "pileon", i.e., with multiple perpetrators coordinated against an individual), to privacy violations (such as stalking, non-consensual sharing of intimate images and "doxing", i.e., publishing private information, such as the target's home address)".[21]

[18] The New York Times, "The Metaverse's Dark Side: Here Come Harassment and Assault", https://www.nytimes.com/2021/12/30/technology/metaverse-harassment-assaults.html?searchResultPosition=2.
There is not one definition of "metaverse", however, it can be understood as "the convergence of two ideas that have been around for many years: virtual reality and a digital second life", see The New York Times, "What's all the hype about the metaverse?" January 18, 2022, https://www.nytimes.com/2022/01/18/technology/personaltech/metaverse-gaming-definition.html.
[19] IWMF, "Attacks and Harassment: the Impact on Female Journalists and Their Work", https://www.iwmf.org/attacks-and-harassment.
[20] The Washington Post, "The only guide to Gamergate you will ever need to read", Caitlin Dewey, October 14, 2014, https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read.
[21] OSCE RFoM, SOFJO Resource Guide, https://www.osce.org/files/f/documents/2/9/468861_0.pdf.

Over the recent years, scores of actors, including international organizations or intergovernmental institutions such as the OSCE RFoM, journalists, newsrooms, media owners, civil society organizations, academia and others have engaged and coordinated to research this no-longer-new frontline, identifying ways to address these attacks and their impact on plurality, media freedom and democracy.

And yet, despite multi-stakeholder processes, a wide range of detailed digital security and anti-harassment manuals,[22] and commitments,[23] the threats continue on an unprecedented level.[24]

According to a recent UNESCO report on trends in the safety of journalists, "seven out of ten women journalists who participated in the global survey reported experiencing online violence during their work, in some cases spurring self-censorship".[25]

THE WAY FORWARD

Recommendations to improve the online safety and digital security of journalists, developed over the course of the last decade, focus on tools and individual measures for those targeted online. Whether through on-going conversations with newsrooms to develop better preventive and protective protocols and/or through guidelines for coordination meetings with State and non-State actors these efforts and measures are important. But alone they seem to be no longer enough. In order to effectively and sustainably improve online safety and security in a gender-responsive way, genuine commitments and clear sets of principles are needed that also provide for specific accountability mechanisms when stakeholders fail to deliver on their commitments. This should include companies (Google, Meta, and others) and platforms (Facebook, Twitter, Telegram), as well as national governments.

Moreover, the international community must acknowledge that the types of attacks - including gender-based violence - taking place online, targeting media practitioners and newsrooms, are not taking place in a vacuum but in combination with policies, decisions, and offline actions that affect them in their daily life and work. This underlines the need for comprehensive responses - encouraging targeted communities to resort to a set of digital security tools, or introducing online harassment measures does not solve the issue of threats and silencing attempts against journalists at heart, it only places a temporary bandage.

---

[22] PEN America, Online Harassment Field Manual, https://onlineharassmentfieldmanual.pen.org/prepare-for-online-harassment.
[23] OSCE Ministerial Council Decision No. 3/18 on the Safety of Journalists, adopted in December 2018, https://www.osce.org/files/mcdec0003%20safety%20of%20journalists%20en.pdf.
[24] UNESCO, ICFJ, "Online violence against women journalists: a global snapshot of incidence and impacts", https://unesdoc.unesco.org/ark:/48223/pf0000375136; IWMF, "Attacks and Harassment: the Impact on Female Journalists and Their Work", https://www.iwmf.org/attacks-and-harassment.
[25] UNESCO, "Threats that silence: trends in the safety of journalists", 2021/2022, https://unesdoc.unesco.org/ark:/48223/pf0000379589/PDF/379589eng.pdf.multi.

Furthermore, while there exists a common understanding that there are various forms of digital threats targeting journalists, developing new tools, and recommendations based on existing threat models is also an outdated approach. For example, existing threat models do not factor in the use of intrusive surveillance software like Pegasus. Neither do these models focus on the role of companies and platforms as an extension of censorship. Numerous examples suggest how companies' lack of understanding of political contexts, deployment of automated response bots, unwillingness or lack of interest in addressing threats faced by smaller, less-known media players, breeds more ground for targeted harassment. All too often, such tools are also not gender-responsive. It is therefore important to keep up with the fast-changing technical development of digital surveillance and harassment methods. Focusing on identifying responses to previously prevalent threats alone, means that the international community lags behind malicious actors whose total disregard for international human rights norms and standards is as rampant as ever.[26]

As long as international commitments remain tame, and violations trigger no consequences, no matter how strong the journalists' passwords are, or how aware journalists are of encryption, secure communication, and website protection, media practitioners in the OSCE participating States and beyond will continue to be easy targets and face risks in online spaces. Unless there is a common agreement to acknowledge existing gaps and align demands as stakeholders, harms done online will continue to hamper plurality, media freedom and hinder the safety of journalists online and offline with a disproportionate impact on women and marginalized voices.

THE THREAT OF DIGITAL AUTHORITARIANISM

Acknowledging the need for comprehensive responses, this paper assesses if and how specific digital security measures designed to protect from attacks such as hacking or DDoS can be useful tools in protecting women journalists from online harassment, hate, and other attacks that threaten or jeopardize their online safety. At the same time, the paper recognizes that digital security measures and response mechanisms must take into account trends in digital authoritarianism worldwide.

The term digital authoritarianism means "the use of digital information technology by authoritarian regimes to surveil, repress, and manipulate domestic and foreign populations".[27] In the introductory remarks to the 2018 Freedom on the Net findings, Adrian Shahbaz from Freedom House wrote, "disinformation and propaganda disseminated online pave poisoned the public sphere. The unbridled collection of personal data has broken down traditional notions of privacy. And a cohort of countries is moving toward digital authoritarianism or increasingly deploy authoritative tools, for example by embracing the Chinese model of extensive censorship and

[26] Kenneth Roth, "How Democracy Can Defeat Autocracy", Human Rights Watch, January 14, 2022, https://www.hrw.org/news/2022/01/14/how-democracy-can-defeat-autocracy.
[27] Alina Polyakova, Chris Meserole, Brookings Institute, Exporting Digital Authoritarianism, https://www.brookings.edu/research/exporting-digital-authoritarianism/

automated surveillance systems".[28] According to the most recent Freedom on the Net report, prepared by Freedom House, global internet freedom has declined for an 11th consecutive year.[29] The findings of the report document that in 2021 alone, at least 20 countries suspended internet access, 21 states blocked access to social media platforms and at least 45 countries were suspected of obtaining and deploying sophisticated spyware or data-extraction technology from private vendors.[30]

Overall, global freedom rankings show that "established democracies lack a consistent and collective strategic approach to combat authoritarian use of digital and online space, even as they often preserve and promote advantageous elements of technology. As a result, concrete actions have not been taken to stem or reverse the pernicious trends of digital authoritarianism," argues Erol Yayboke in his Center for Strategic and International Studies brief.[31] Overall, authoritarian trends often go hand in hand with increasing attacks against women, journalists, and civil society. Digital authoritarian tools are often deployed in an attempt to consolidate power and control, and to restrict the diversity of voices. In doing so, they regularly instrumentalize existing inequalities, and disproportionally curtail the safety of women and marginalized groups.

EXISTING DIGITAL SECURITY SOLUTIONS FOR JOURNALISTS AND MEDIA PRACTITIONERS AT LARGE

At the time of the revelations by the NSA contractor Edward Snowden in 2013, digital security tools that journalists often rely on today, such as end-to-end encrypted communication, were rare, difficult to use and inaccessible.[32] If anything, there were not many resources available at all. Today, digital security has come into focus as a precondition to the exercise of online rights, and the landscape of available tools has changed significantly, with some as a direct response to gender-based online abuse. Scores of organizations – the Rory Peck Foundation, Reporters Without Borders, PEN America, Frontline Defenders, Freedom of the Press Foundation, Free Press Unlimited, Committee to Protect Journalists, the Electronic Frontier Foundation and many others – as well as media outlets themselves have developed, compiled and published lists of resources, useful and comprehensive manuals and valuable information on digital security as well as safety tools, measures and mechanisms. The tools themselves, such as end-to-end encryption have also become less sophisticated and thus more accessible, offering their users quick and easy access to secure communication online. Many of these applications, software and manuals are thus used not

[28] https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism
[29] Freedom House, "The Global Drive to Control Big Tech", https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech.
[30] Ibid.
[31] Erol Yayboke, Center for Strategic and International Studies, Promote and Build: A Strategic Approach to Digital Authoritarianism, https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism
[32] Freedom of the Press, "How Snowden Has Changed Journalism and Privacy Five Years Later," https://freedom.press/news/how-snowden-has-changed-journalism-and-privacy-five-years-later.

only by journalists and media practitioners but across civil society organizations as well as the general public.

In addition to readily available manuals, international media and press freedom advocates have dissected the common digital threats journalists face today. While there is no clear definition of online safety or digital security, both should be understood as broad concepts.

Overall, there has been a progress when it comes to awareness on digital security and online safety, including on the distinct risks faced by women. Ela Stapley, the digital security advisor at the International Women's Media Foundation, who was interviewed for this report, notes the overall change in conversation and awareness around threats faced online and importance of digital safety. She emphasized that social media platforms too have taken notice of the scale of online harassment and threats online, as have some governments such as those who have spearheaded initiatives such as the Freedom Online Coalition that also focuses on safety of journalists (and specifically women journalists) online vis-à-vis platforms.[33]

However, notes Stapley, in the case of social media platforms, while they acknowledge that online abuse is a tech issue, they have failed to invest in (human) resources to effectively tackle harassment on their platforms, especially from a gender perspective. Instead, they seem to rely on civil society organizations and international non-governmental organizations such as Access Now,[34] Reporters Without Borders, Committee to Protect Journalists and others as "gatekeepers" and supporting actors between targeted users and the platforms. As threats escalate and the number of journalists targeted online grows, so does the burden placed on third party organizations that intervene on behalf of targeted journalists. Thus, the necessity for platforms to do more when journalists face online abuse becomes even more evident.

The overreliance of platforms on machine-learning algorithms poses several human rights concerns, in particular as these algorithms are shielded from any external review, negating the principles of transparency and accountability.[35] As a report by the Electronic Frontier Foundation puts it: "Civil Society, and governments have been denied access to the training data or basic assumptions driving the algorithms, and there has never been any sort of third-party audit of such technology".[36]

[33] Interview with Ela Stapley.
[34] The author of the report is on the board of Access Now.
[35] OSCE RFoM, "Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual", https://www.osce.org/files/f/documents/8/f/510332_0.pdf.
[36] Jillian C. York, "Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content the Net", May 30, 2019, https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content.

The inherent belief that AI systems are less biased and scale better than human beings, is inaccurate, in particular when it is not adjusted to languages other than English.[37] Technology is also not gender neutral. It is crucial that platforms engage with individuals with contextual knowledge and an understanding of political and social contexts as well as local languages, and keep humans in the loop. Otherwise, cases where news reporting is taken down from social media platforms, for example based on takedown requests by state actors for alleged problematic or pornographic content, will continue to exist and the states' silencing efforts will continue to succeed, especially if it is not possible to receive anything but an automated response.

In some contexts, platforms like Facebook rely on lists of allegedly dangerous individuals and organizations in their content moderation practices. This results in news reports, mentioning names on the list, flagged and potentially blocked from social media platforms because of their reporting on organizations designated as terrorists.[38] This approach to content moderation fails to distinguish between objective reporting and actual propaganda. There are currently no safeguarding mechanisms to prevent undue account suspension. Moreover, journalists who are independent or who do not work for bigger, well-known media organizations often do not receive responses from platforms and are not given the same attention in terms of user safety.

On the other side of moderation, and equally problematic from a media freedom perspective, is a lack of action or refusal to take down content that is actually posing a threat to the safety of journalists. This disproportionally affects women in the public sphere, and even more so those with intersecting identities. For example, when news outlets report to platforms that their journalists are being doxed, the platform does not remove the reported posts because the platforms do not consider (based on automated decision) the posts in violation of their community standards. As such, in cases that pose high risks particularly to women journalists it is essential to ensure human review and that decisions are made by an individual aware of the political context and the kind of targeting the journalists are facing.

Across the board, regardless of users' professional affiliation, whether they are journalists, activists, or human rights defenders, a reoccurring problem is that it is virtually impossible to reach and communicate with real individuals at these platforms. All too often, it takes an intervention by a third party that has direct contacts at these platforms and companies to address an issue at stake and resolve a specific threat problem.

According to Courtney Radsch, former Advocacy Director at the Committee to Protect Journalists and currently visiting scholar at the Center for Media at Risk at the Annenberg School for

[37] The Conversation, "Beyond a technical bug: Biased algorithms and moderation are censoring activists on social media", May 16, 2021, https://theconversation.com/beyond-a-technical-bug-biased-algorithms-and-moderation-are-censoring-activists-on-social-media-160669.
[38] For the Facebook list of Dangerous Individuals and Organizations, see https://theintercept.com/document/2021/10/12/facebook-dangerous-individuals-and-organizations-list-reproduced-snapshot.

Communication, platforms must take a more proactive role in identifying and labeling the type of threats individuals can face on their services. Depending on whether it is a coordinated or rather individually targeted attack, platforms should offer different mechanisms to mitigate the attack and assist the targeted user. Focusing on coordinated attacks can be helpful, states Radsch, because this is where the AI and machine learning deployed by platforms can support mitigating against inauthentic behavior and coordinated harassment online.[39]

Some practices of authoritarian-leaning governments leveraging social media platforms to restrict the free flow of information, however, call for even stronger measures. Some countries restrict online content by sending government requests to platforms to remove content or turn over information about users, writes Ron Deibert, professor of political science and the founder of Citizen Lab at the Munk School of Global Affairs at the University of Toronto.[40]

Media practitioners face additional challenges in countries where the national governments exhibit hostile behavior towards independent journalists. Often combined with offline measures such as defamation suits, strategic litigation against public participation (SLAPPs), arrests, and traditional forms of intimidation, these governments also deploy a range of online tools, including the use of troll armies, account compromise, DDoS attacks, and the man-in-the-middle attacks[41] against news websites, spear-phishing[42], or excessive content removal requests as well as other forms of gender-based violence.

As digital threats against media practitioners and particular women journalists continue to grow, there has been a shift of digital security awareness and measures within newsrooms too. A difference in approaches can be seen depending on whether the newsrooms operate within challenging environments or jurisdictions with stronger safety nets.

While newsrooms have taken stock of such attacks and other forms of online harassment against journalists working with them, or targeting themselves as outlets, consistency in addressing such attacks is often missing. Consistency, however, would be necessary for digital security measures to become a part of the newsroom culture. Courtney Radsch notes that while digital security practices are becoming more common, until they are fully integrated with newsrooms' onboarding and offboarding practices and human resources, and are adopted across the board of services (including social media presence, everyday practices, etc.) for everyone within the newsroom,

[39] Interview with Courtney Radsch.
[40] Excerpt from Chapter 17, "Digital threats against journalists", Journalism After Snowden, https://gijn.org/2017/06/13/journalism-after-snowden-the-growing-digital-threat-to-the-press.
[41] This refers to a cyberattack where communication between two parties is altered or manufactured while the parties believe they have private and direct connection.
[42] This refers to an electronic communication scam with the intention to gain access to an individual's account or impersonate a specific individual.

there will be no effective or sustainable change. It is a process that will take time, notes Radsch, emphasizing again the need for consistency.[43]

IMPACT ON INDIVIDUAL JOURNALISTS

Today, journalists live and operate in an environment where not only they are facing a more sophisticated technology targeting them personally, their work, colleagues, and sources, but also at a time of targeted, often gender-based and government-sponsored, deliberate harassment, and disinformation campaigns. Digital authoritarianism is reaching more corners of the world, while established democracies seem to lack effective coping strategies. As a result, threats against journalists remain and the nature of the attacks become more brutal and prolific, while also continue to "erode the foundations of journalism by chipping away at journalists' resolve to provide independent, critical reporting of crucial issues."[44]

Moreover, while the list of potential risks journalists face continues to change and evolve, so does the landscape of actors, technology, and platforms. It is thus becoming increasingly harder for journalists to stay up to date with rapid changes of their digital security and online safety environment, especially in the face of attacks prompted by technologies such as Pegasus and other intrusive surveillance. This is particularly true when the attacker has resources and technology available that journalists and the general public do not have.

While threats continuously change and become more organized and sophisticated, the response mechanisms do not. This is not necessarily because all existing tools are outdated but because the international community and responses lag behind. Digital security is ever-changing. One tool may work once but may not work the next time in a different context.

TOWARDS HOLISTIC APPROACHES MOVING FORWARD

As Tim Berners-Lee, the founder of the World Wide Web wrote in his 2018 opinion piece for The New York Times, "just as the web was built by millions of people collaborating around the world, its future relies on our collective ability to make it a better tool for everyone".[45]

As digital security practices cannot alone solve the issue of online abuse in its various forms, there is a pressing need to engage a whole-of-society approach with all relevant stakeholders, from journalists, newsrooms, to international organizations and governments.

[43] Interview with Courtney Radsch.
[44] UNESCO and ICFJ, "Online violence against women journalists: a global snapshot of incidence and impacts", https://unesdoc.unesco.org/ark:/48223/pf0000375136.
[45] Tim Berners-Lee, "How to Save the Web", The New York Times, December 6, 2018, https://www.nytimes.com/2018/12/06/opinion/tim-berners-lee-saving-the-internet.html.

Effective training, not only manuals: Experts interviewed for this report agree that there is no need to develop yet another manual on digital safety for journalists. Instead, updating existing manuals periodically and incorporating new case studies and threat models as well as a strong gender and intersectional perspective would be more effective. This should be developed in partnership with journalists who have experienced digital threats and online harassment and who can shape the conversation around new risk assessments and threat models. A focus should be put on trainings. A particular added value would be to also bring in journalists as trainers, as they can shape the narrative around the importance of getting the story, i.e., how adopting security practices can help investigations and researching a story safely.

Moreover, due to the multi-faceted threats against journalists, there is a need for a comprehensive training approach that includes all aspects of safety (physical, digital, and psycho-social safety), and is intersectional, addressing the additional risks faced by women journalists and other marginalized communities.[46]

Regional or country-specific trainings can be additionally useful, but require bringing in experts from the region and journalists aware of distinct risks and threats, who can take part in designing and holding these trainings to address existing contextual, language and cultural deficiencies in trainings.

Generally, a more **holistic approach to safety** of journalists is needed, not just when training journalists on digital safety and security, but one that extends to their newsrooms. Many media outlets have already implemented security protocols and strengthened infrastructure and security measures. Further engagement with organizations that provide digital security support and trainers could enable crucial conversations around safety of journalists online, and that these safety measures are followed holistically, starting from onboarding to human resources and IT support policies within newsrooms. A holistic approach requires gender-responsive analyses, strategies and ensuring an intersectional perspective.

Ensuring the basics of digital security for journalists and newsrooms: Just as locking our doors when leaving homes or offices, newsrooms should adapt to using basic digital security tools. The use of encrypted communication, two-step verification, having a digital security helpdesk on call, readiness to remove possible doxing information with the help of available guides should be encouraged if not made mandatory.[47] However, these measures are only one part of a broader set of measures, which must be in place in order to mitigate the risks of digital attacks taking into account the specific and often intersecting threat context in which they take place. No encryption, no matter how sophisticated it is, will fully protect journalists from the risks outlined above. In an environment where information controls used to target journalists, it is essential to change

[46] https://www.ecpmf.eu/wp-content/uploads/2021/04/DG-CNECT-Recommendations-on-Safety-of-Journalists-3.pdf
[47] Feminist Frequency, "A guide to protecting yourself from online harassment", https://onlinesafety.feministfrequency.com/en.

communication practices through behavioral changes as well.[48] Any regulation or policy regarding communication surveillance or content control needs to be in line with international human rights standards.

Education and awareness: As one expert interviewed for this report pointed out, online harassment is a societal issue, and thus requires societal change, which requires campaigning with newsrooms, platforms and companies, as well as governments to introduce holistic concepts around the importance of online safety into national legislation.

Changing the culture in newsrooms: In addition to protocols for newsrooms to support journalists targeted with online harassment,[49] creating a gender-responsive standard in news industry on digital security and safety would be useful. Any adopted mechanism, however, must be fluid, holistic, regularly updated and followed through to be effective.

Role of platforms: While all major platforms' terms and services or community standards prohibit online attacks and harassment, as well as impersonation and other digital security threats, the success rate, effectiveness and enforceability of such rules have repeatedly been questioned.[50] In July 2021, Facebook, Google, TikTok, and Twitter committed themselves to better tackle the abuse of women on their platforms.[51] While the list of commitments such as "offering more granular settings (who can see, share, comment or reply to posts); offering users the ability to track and manage reports; enabling greater capacity to address context and or language, etc," is impressive, its success rate, its usability across all countries and different languages is yet to be seen and evaluated. To be beneficial, the procedures for handling and resolving complaints must be transparent, easy-to-use and effective. Any rules for digital platforms are only effective if they are properly enforced.

The Santa Clara Principles on Transparency and Accountability in Content Moderation,[52] launched in 2018 and iterated in 2021, for example, provide guidance on human rights-friendly content moderation. Since their launch in 2018, major companies such as Apple, Facebook, Google, Twitter have endorsed the guidelines. The iteration of the document introduced additional foundational and operational principles. While the former includes "overarching, and cross-cutting principles that should be taken into account by all companies", the latter looks at more specifics "regarding precisely what information is needed to ensure meaningful transparency and

[48] Global Investigative Journalism Network, "Journalism after Snowden: The growing digital threat", Ron Deibert, https://gijn.org/2017/06/13/journalism-after-snowden-the-growing-digital-threat-to-the-press.
[49] International Press Institute, "Protocol for Newsrooms to support journalists targeted with online harassment", https://newsrooms-ontheline.ipi.media/wp-content/uploads/2020/02/IPI_newsrooms_protocol_address_online_harassment_ok_022020.pdf.
[50] OSCE RFoM, "Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual", https://www.osce.org/files/f/documents/8/f/510332_0.pdf.
[51] Web Foundation, https://webfoundation.org/2021/07/generation-equality-commitments.
[52] Santa Clara Principles on Transparency and Accountability in Content Moderation, https://santaclaraprinciples.org.

accountability". The iteration also expanded the "scope of where transparency is required with respect to what is considered 'content' and 'action' by a company."

To ensure the use of automation in content moderation and curation does not infringe on human rights, the OSCE RFoM recently published a comprehensive Policy Manual, putting a spotlight on artificial intelligence and freedom of expression (SAIFE). It includes a chapter on hate speech and other harmful content, providing guidance on how states can ensure the principles of transparency, accountability and public oversight are materialized, and that free speech safeguards are provided for the use of AI.[53]

OSCE follow-up mechanisms and taking stock: Since the inception of the SOFJO project in 2015, the OSCE RFoM has raised awareness on the individual and societal impact of online harassment and abuse against women journalists, and provided guidance on how to ensure more online safety. This and the Office's general work on safety, including online, contributed to the Ministerial Council Decision No. 3/18 on the Safety of Journalists. In this Decision, participating States recognize that the work of journalists can put them at risk of violence, intimidation and harassment, which can deter them from carrying out their work or lead to self-censorship, thereby having a negative effect on media freedom and media plurality. The Decision further recognizes that targeted campaigns undermining the work of journalists are increasing, eroding public trust and confidence in the credibility of journalism, which in turn can increase the risk of threats and violence. It also recognizes that safety entails physical, legal, political, technological, economic and other aspects, and outlines threats such as hacking or arbitrary surveillance, with States committing themselves to refrain from unlawful or arbitrary interference with encryption, surveillance and other attacks on digital security. Also, participating States recognize the distinct risks faced by women journalists and commit themselves to add a gender perspective to their safety of journalists' efforts.[54] The SOFJO Resource Guide released in 2020 outlines specific proposed actions for stakeholders across the board, to provide guidance on *who* can do *what* – and *how* – to improve the online safety of women journalists. However, visible changes are still outstanding, and more efforts are needed to close the implementation gap. This year marks the 25th anniversary of the OSCE RFoM mandate. One initiative the institution could employ with regard to enhancing media freedom across the OSCE region, is to acknowledge gaps, and to issue a unified communique asserting the institution's commitment to provide rapid responses to violations of the organization's key principles with regard to media freedom and freedom of expression, and to take a more pro-active role in bringing together all stakeholders to outline concrete guidance, deadlines and accountability mechanisms. This should include a stronger focus on digital security, which currently is missing in various OSCE discussions and initiatives.

[53] OSCE RFoM, "Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual", https://www.osce.org/files/f/documents/8/f/510332_0.pdf.
[54] OSCE Ministerial Council Decision on the Safety of Journalists, Milan 2018, https://www.osce.org/chairmanship/406538.

Legal approach: It would be useful to map existing legal remedies across OSCE participating States that can be used in cases of digital threats as accountability measures. This mapping should include what, if any, punitive measures exist to address digital threats, let alone from a gendered perspective. It should also include an assessment of the measures as monitored by independent observers, including on the transparency of decisions and steps taken in this context.

Decentralized web: Tim Berners-Lee underlined that "the web needs radical intervention from all those who have power over its future: governments that can legislate and regulate; companies that design products; civil society groups and activists who hold the powerful to account; and every single web user who interacts with others online".[55] He re-introduces the concept of a decentralized web, where the power is returned to the people and users, away from governments, companies and platforms who are currently benefiting from the internet's centralized nature. This new decentralized structure would entail more built-in anonymity, with the caveat that this also entails chances of empowering malicious behavior – trolls just as harassers. Although this new impetus for a decentralized, more anonymous structure is controversial to some degree, there are many actors, developers, and activists currently working towards this idea. In light of failures in pushing the narrative around transparency and accountability online, this may be one way to move forward. It may require restructuring our approaches and existing mechanisms but it is nevertheless worth considering, as it may be the future.

[55] Tim Berners-Lee, "I Invented the World Wide Web, Here's How We Can Fix It", The New York Times, November 24, 2019, https://www.nytimes.com/2019/11/24/opinion/world-wide-web.html.

ABOUT THE AUTHOR

Arzu Geybulla is an Azerbaijani columnist and writer with a special focus on digital authoritarianism. Arzu has written for Al Jazeera, Eurasianet, CODA, Open Democracy, and Radio Free Europe, with a byline on CNN International. She is also a regional editor for South Caucasus and Turkey at Global Voices. In 2019, Arzu launched Azerbaijan Internet Watch, a platform that documents, and monitors information controls in Azerbaijan. Since 2015, she has been involved in various projects focusing on the safety of women journalists online. Arzu is based in Istanbul from where she continues her journalism work as well as her engagement in projects that continue to focus on the safety of women journalists online, platform accountability, and transparency.