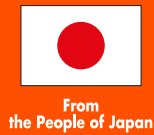


March 2025

# CASES AND RESEARCH

**FROM THE WAR IN UKRAINE**

A supporting document for the  
Guide for Risk Management  
in the context of emergencies,  
armed conflicts and crises.



## Authors

Pavlo Belousov (NGO Internews Ukraine)  
Nynne Storm Refsing (IMS)

## Working group members

Yaroslav Yurchyshyn  
Oleksandr Burmahin  
Ganna Krasnostup  
Liza Kuzmenko  
Alisa Malytska  
Yaroslava Dyo  
Igor Rozkladai  
Nataliia Tkachuk  
Andrii Myroshnichenko  
Iryna Zemliana  
Maksum Dvorovyi  
Alona Romaniuk  
Andrii Shevchenko  
Oksana Moroz  
Olha Yurkova  
Mykola Typusiak  
Anton Melnyk  
Maksym Onopriienko  
Maksym Savanevsky  
Olena Dub  
Dmytro Sholomko  
Mariya Frey  
Anastasia Ostrovska  
Valentyna Aleksandrova

## Other contributors

Maria Mingo  
Elodie Vialle  
IMS' High-Level Group of Experts for Resilience Building in Eastern Europe (HLEG)

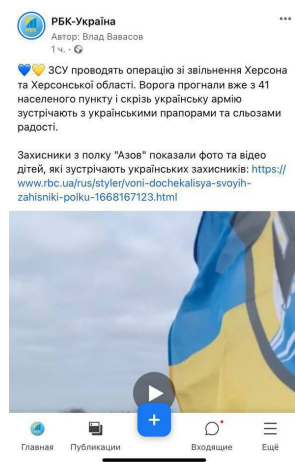
*This publication was drafted within the framework of a project implemented by IMS (International Media Support) and NGO Internews Ukraine in partnership with UNESCO and with the support of Japan. The authors are responsible for the selection and presentation of the facts contained in this publication. The views expressed are solely those of the authors and do not necessarily reflect the position of UNESCO or Japan. The project builds on UNESCO's Guidelines for the Governance of Digital Platforms from 2023.*

## Cases and research

The following cases and research exemplify the risks identified in the Guide for Risk Management in the context of emergencies, armed conflicts and crises ('the Guide'). They are accompanied by examples of responses from social media platform companies to challenges associated with the full-scale invasion of Ukraine.

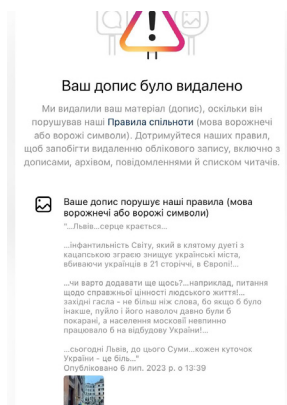
The cases and research contain risk-specific recommendations to complement those outlined in the Guide for social media platform companies to respond to hate speech and disinformation, including disinformation and hate speech targeting and affecting women, minorities and marginalised groups in contexts of crisis, emergency and armed conflict in the future. They also highlight useful actions taken by companies to adapt to the context of emergency, crisis and armed conflict.

### Case 1: Blocking and/or reducing the reach of war-related content that does not violate community standards.

<b>Date / period:</b>	November 2022
<b>Platform:</b>	Facebook
<b>Description:</b>	On 1 March 2022, the Ukrainian city Kherson was occupied by Russian forces. After 256 days of occupation, on 11 November 2022, the Armed Forces of Ukraine liberated Kherson. In November, a video showing children peacefully meeting Ukrainian soldiers of the Azov regiment in the liberated Kherson was published on the news agency RBC-Ukraine's Facebook page. Following user reports, allegedly from Russian users, Facebook moderation mechanisms subsequently restricted the visibility of this and other content on the page of RBC-Ukraine.
<b>Platform response:</b>	Facebook reduced the reach of content from RBC-Ukraine's page, citing violations of community standards. RBC-Ukraine appealed to Facebook to reverse their decision and encouraged its readers to do the same. On 19 January 2023, Ukrainian Minister of Digital Transformation Mykhailo Fedorov announced that <a href="#">Meta had declared</a> that they will not block content about the Azov Regiment.
<b>Source/ further information:</b>	<p>In an <a href="#">article</a> published on 22 November, RBC-Ukraine stated that it faced content restrictions on Facebook following a complaint about a post mentioning the Azov regiment.</p> <p>The screenshot here shows the Facebook post itself with a video of residents of Kherson greeting the Ukrainian army in liberated territories. <a href="#">Snapshot of the page</a>.</p> 
<b>Case relevance:</b>	In situations like the conflict in Ukraine, where an aggressor employs information warfare tactics, restrictions on war-related content that does not violate community standards can interfere with the fundamental human right to access to information. The right to freedom of expression encompasses the public's ability to seek, receive, and impart information of all kinds, especially during times of crisis. By removing or limiting content that provides critical insights into the realities of the conflict, these actions not only limit plurality and diversity, but can also inadvertently amplify disinformation and hate speech, ultimately undermining efforts to protect the right to truth and informed decision-making. Such content moderation decisions may also contribute to a chilling effect, increasing unjustified fears and confusion among the civilian population, further exacerbating an already precarious situation. Moreover, undermining access to reliable information may contribute to a loss of trust in platforms, potentially


	violating the public's right to participate in informed public discourse. In this context, content moderation must be carefully balanced to prevent further manipulation of information while ensuring the protection of human rights and democratic principles.
<b>Risk-specific recommendations:</b>	<ul style="list-style-type: none"> <li>a) Efforts should be strengthened when it comes to raising awareness and providing training for private users, media, civil society and other relevant stakeholders in community standards and moderation policies.</li> <li>b) Ongoing monitoring should be conducted for both algorithmic and manual decision-making regarding potential over- or under-enforcement of moderation policies. Regular assessments with local stakeholders will help identify issues.</li> <li>c) Transparency in the development and application of content moderation policies should be increased, providing users with access to clear and detailed explanations for removing or reducing the reach of their posts. For example, instead of limiting it to a general phrase like "violates community standards," it should specifically indicate which part of the policy was violated and how the content contradicts the relevant norms.</li> <li>d) Companies should include gender-based violence or similar as a reporting category for users.</li> <li>e) Appeal and review mechanisms in the Ukrainian language should be made as accessible, understandable, and transparent as possible for users.</li> </ul>

## Case 2: Removal of content that documents war crimes.

<b>Date / period:</b>	June 2023
<b>Platform:</b>	Facebook, Instagram, YouTube
<b>Description:</b>	In June 2023, journalist Ihor Zakharenko, while documenting the aftermath of Russian attacks on civilians in Kyiv's suburbs, encountered a problem with his videos being removed from his accounts on Facebook, Instagram, and YouTube. The materials, which contained evidence of war crimes, were automatically deleted shortly after they were uploaded. This occurred due to the operation of algorithms that social networks use to filter content containing scenes of violence. While intended to protect users from harmful content, these algorithms often do not consider the context in which the material was created, leading to the removal of important evidence of human rights violations.
<b>Platform response:</b>	<p>Meta and YouTube stated that content from war zones could remain on the platforms if it was in the public interest. However, in practice, algorithms often delete such materials automatically. Meta noted that they respond to requests from law enforcement agencies worldwide and continue to explore additional opportunities to support international accountability processes, in accordance with legal and privacy obligations.</p> <p>YouTube emphasised that while there are exceptions for content of public interest, the platform is not an archive, and recommended that researchers, human rights defenders, and journalists use other methods to preserve important materials.</p>
<b>Source/ further information:</b>	<p>In an article published in June 2023, <a href="#">BBC Ukraine</a> provided examples of the reasoning and ways in which social networks delete evidence of war crimes in Ukraine. The material is shown via video reportage on <a href="#">YouTube</a>.</p> <p>The screenshot here shows similar messaging in another case where an Instagram <a href="#">post</a> of a Ukrainian journalist which documented war crimes was removed as it was deemed by platform moderation mechanisms to contain hate speech.</p> 
<b>Case relevance:</b>	In the context of the ongoing conflict in Ukraine, documenting war crimes committed by Russian and Belarusian forces is a fundamental human rights issue. The right to justice, truth, reparation and non-repetition requires that such atrocities be investigated and the perpetrators held accountable. Access to accurate and timely documentation of these crimes is crucial not only for the victims but also for Ukrainian authorities and the international community, as they seek to safeguard Ukraine's sovereignty and defend against invasion. Whilst platforms have their own processes with law enforcement authorities, these are not always sufficient and effective, consequently leading to authorities turning to civil society groups for access to information and removed content.

<b>Risk-specific recommendations:</b>	<ul style="list-style-type: none"> <li>a) Collaboration with local and international organisations should be initiated to ensure that content documenting war crimes removed by the company is preserved for three to five years.</li> <li>b) Retained content documenting war crimes posted by users which is removed from the company platform(s) should be stored outside of Ukraine in a low-risk context in accordance with international standards on privacy and data protection.</li> <li>c) Collaboration with law enforcement authorities and civil society organisations with experience in archiving and data anonymisation should be implemented to improve and disclose identification parameters of critical human rights content for retention, as well as content triaging processes for sharing with law enforcement authorities.</li> <li>d) Users should be notified if their content is retained by companies.</li> <li>e) Relevant local stakeholders should be notified of the existence of archived materials relevant for processes to ensure accountability and justice for war crimes.</li> <li>f) Special attention should also be provided to crimes against women, minorities and marginalised groups, and conflict-related sexual violence and collaborations initiated with expert rights organisations.</li> </ul>
---------------------------------------	---

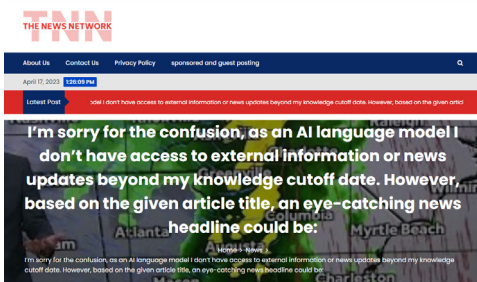
### Case 3: The use of bots and fake accounts to spread disinformation and hate speech.

<b>Date / period:</b>	November 2023
<b>Platform:</b>	TikTok
<b>Description:</b>	<p>In November 2023, numerous accounts on TikTok had videos which promised non-existent monetary payments ranging from 9,000 to 50,000 UAH (roughly USD 200-1,200) for residents remaining in Ukraine after the start of the full-scale invasion and urged users to follow links that led users to subscribe to anonymous bot Telegram channels. These channels were filled with disinformation, misleading polls on political topics, and unreliable information about the war in Ukraine, for example, false government decisions.</p>
<b>Platform response:</b>	At the time of the study, there has been no response from TikTok regarding the identified accounts.
<b>Source/ further information:</b>	<p>The Institute of Mass Information (IMI) presented their research on the spread of disinformation in Ukrainian on TikTok in an <a href="#">article</a> published in January 2024. It documents how fake TikTok channels are used to direct users to anonymous Telegram channels.</p> <p>Most of the identified fake videos on TikTok are compiled from clips of news programmes from popular Ukrainian TV channels. The screenshot here shows false information about international aid to Ukraine (7000 UAH from the US) as if shown on the popular TV channel 1+1. <a href="#">Page snapshot</a>.</p> 
<b>Case relevance:</b>	<p>Exploiting the fears and insecurities of Ukrainians—in this case related to financial hardship—through disinformation and scamming is a violation of the public's right to accurate and truthful information. The spread of such harmful content not only undermines individuals' ability to make informed decisions but also jeopardises the mental well-being and security of Ukrainians, including vulnerable groups such as young people and socio-economically disadvantaged groups. Targeted disinformation can contribute to the erosion of trust in vital state institutions and authorities, thereby weakening social cohesion and undermining the right to live in a peaceful and secure environment.</p> <p>Disinformation that plays on people's fears exploits their vulnerability and amplifies feelings of anxiety, panic, and confusion, which can lead to increased societal tensions. In this context, the responsibility to protect individuals from harmful content falls on both state authorities and private platforms. Platforms like TikTok must ensure that they effectively protect the human rights of users by preventing the spread of such harmful material, especially to those who are most at risk of being misled. At the same time, governments must ensure that policies and actions protect the right to information, promote media literacy, and build resilience among vulnerable communities to combat disinformation and prevent its negative impact on society.</p> <p>Though the issue of bots in this case was located on Telegram, significant challenges of inauthentic activities and bots and vulnerabilities in detection thereof have been detected on <a href="#">all major social media platforms</a>.</p>

	<p>Cases have been documented of bots and trolls that comment and like posts which support disinformation narratives or attack reliable content and data about Russian aggression, and they are successful in creating an illusion of mass support for certain views. Disinformation tends to build on harmful stereotypes which is often highly gendered and sexualised in its character. It seeks to fuel societal polarisation and dispute particularly targeting and impacting women, minorities and marginalised groups. Moreover, fake accounts posing as Ukrainian military personnel or volunteers spread manipulative and unreliable information about Ukrainian army losses and false news about humanitarian catastrophes. Such actions sow panic and despair, and undermine trust in Ukrainian institutions.</p> <p>This manipulation not only compromises individuals' right to security but also infringes on their dignity and freedom from harm. Both governments and platforms have a responsibility to protect individuals by preventing the spread of disinformation, promoting media literacy, and safeguarding the right to truthful information, particularly during times of crisis.</p>
<b>Risk-specific recommendations:</b>	<ul style="list-style-type: none"> <li>a) Resources should be prioritised to identify, moderate and block fake and automated accounts, pages and groups that post and share disinformation and hate speech with particular attention to gendered disinformation and discriminatory content.</li> <li>b) Efficient early warning systems and escalation channels should be established with local stakeholders, for example, trusted fact-checkers, media, gender experts and civil society organisations, for emergency cases of inauthentic mass activities promoting particularly dangerous disinformation and hate speech narratives, including narratives fuelling prejudice and societal polarisation and targeting women, minorities and marginalised groups.</li> <li>c) Regular monitoring and evaluation should be conducted of algorithmic and manual decision-making in company transparency reports. Key takeaways should be shared and discussed with local stakeholders.</li> <li>d) Cooperation should be established with national and local stakeholders who are well-placed and have the professional capacity to produce reliable, fact-checked information and counter harmful and misleading content. These stakeholders should also have access to and be able to provide crisis and frontline insights to improve data sets that algorithmic moderation is built upon.</li> <li>e) Fact-checked content countering disinformation narratives should be identified and promoted in the immediate aftermath of spikes in mass activities promoting harmful narratives and in proportion to these activities. This should be done with specific attention to disinformation compromising human rights including women's rights and anti-discrimination.</li> <li>f) Fact-checked content in the public interest related to the war – for example, related to security and humanitarian assistance – should be identified and promoted in collaboration with crisis experts and local stakeholders.</li> <li>g) Special attention should be paid to authentic accounts with a long history that may be "rented out" to spread disinformation. Such cases should be identified through monitoring sudden increases in suspicious activity and tracking atypical account behaviour.</li> </ul>

#### Case 4: Presence of false, misleading, and malicious content, including AI-generated content and content targeting women, minorities and marginalised groups.


<b>Date / period:</b>	May 2023
<b>Platform:</b>	Google
<b>Description:</b>	In May 2023, a network of 49 news websites containing artificially generated content was discovered. The sites, operating in various languages, posted articles spreading disinformation and promoting Russia's war against Ukraine. One of the sites, TNewsNetwork, published a false story about the death of thousands of soldiers in the war, based solely on a YouTube video that contained no verified facts.
<b>Platform response:</b>	Following an investigation, Google removed advertisements from specific pages on these sites. In cases of systematic violations, advertising was completely blocked for the entire site. However, the presence of AI-generated content is not automatically a violation of advertising policies. But if automation, including AI, is used to manipulate ranking in search results, then it is considered a violation of spam policies.
<b>Source/ further information:</b>	Articles on <a href="#">Bloomberg</a> and <a href="#">NewsGuard</a> from spring 2023 document how AI tools are used to generate content at so-called content farms: low-quality websites with large quantities of AI-generated content that is not fact-checked and often spread false information.

	<p>The screenshot below shows an AI-generated headline that appeared on TNewsNetwork.com, an anonymously run news site registered in February 2023.</p> 
<b>Case relevance:</b>	<p>Such sites as the example of the case often meet formal requirements of social media platforms, allowing them to participate in advertising programmes and generate profits while spreading disinformation.</p> <p>Automation enables the rapid spread of disinformation, which is often used to fuel distrust, societal polarisation, and harmful stereotypes and gender-based violence. This includes content that targets women, minorities, and marginalised groups, spreading false narratives about the Ukrainian military and institutions. Examples have been documented of deepfake videos of Ukrainian politicians causing panic and undermining trust in democratic institutions.</p> <p>Social media platforms can amplify gendered disinformation and hate speech, particularly against women in public roles, such as politicians, journalists, human rights defenders and LGBTQIA+ activists. Such information campaigns are aimed at undermining trust in women leaders by spreading false information that questions their competence, moral character, or loyalty to the state. This type of online violence worsens gender inequality, violates the right to freedom of expression, and undermines the right of individuals, especially women, minorities and marginalised groups, to participate in public life free from discrimination and harm.</p>
<b>Risk-specific recommendations:</b>	<ol style="list-style-type: none"> <li>Mechanisms should be introduced for temporary restriction or controlled reduction of reach for content that shows signs of false, misleading, or malicious information, including that generated by artificial intelligence and with specific attention given to gendered disinformation and discriminatory content.</li> <li>Efficient early warning systems and escalation channels should be established for emergency cases that could impact the physical safety of individuals.</li> <li>Priority should be given to allocating sufficient financial and human resources both for detecting audio, visual, and audiovisual content generated by artificial intelligence, and for developing new technologies that improve the accuracy of such identification.</li> <li>Labelling of artificially produced text, audio, visual, and audio-visual content should be introduced.</li> </ol>

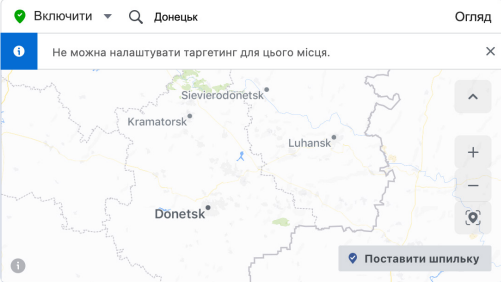
**Case 5: Moderation policies and practices lacking consideration of contextual linguistic, social, political, historical, and cultural understanding, including about gender, minorities and marginalised groups.**

<b>Date / period:</b>	January 2024
<b>Platform:</b>	Instagram
<b>Description:</b>	Between October and December 2023, the Instagram pages of "Memorial", an account collecting stories about <a href="#">fallen and killed soldiers</a> and <a href="#">civilians</a> , were blocked four times, and in December 2023, the accounts were removed from Instagram by Meta. Since March 2022, these accounts had been publishing stories about Ukrainian soldiers lost during the war. The deletion resulted in the loss of access to 5,440 detailed accounts created to preserve the memory of the soldiers.
<b>Platform response:</b>	After a wave of user support for the unblocking of the account, <a href="#">Meta restored</a> the Memorial Instagram accounts in January 2024.
<b>Source/ further information:</b>	The independent media outlet Detector Media, an expert organisation in disinformation and propaganda, in an <a href="#">article</a> published in January 2024, documented the restored Instagram Memorial accounts after protests by Ukrainian users. <a href="#">Page snapshot</a> .



	<p>The screenshot below shows a <a href="#">Facebook</a> post by the coordinator of the "Memorial" project who asks the community to sign up for their new Instagram page after initial page deletion.</p> 
<b>Case relevance:</b>	<p>Blocking popular accounts with verified information deprives Ukrainians of their right to express grief and honour fallen soldiers and civilians, limiting public awareness of the war's toll and weakening support for affected families. Such actions also erode trust in social media platforms, especially when users face content removal or account suspension for criticising propaganda or making local, culturally relevant jokes. This restricts Ukrainians' freedom of expression, hindering their ability to participate in important public debates and share their experiences. It's especially problematic when platforms fail to recognise the sensitive context of local expressions, risking discrimination against women, minorities, and marginalised groups.</p> <p>Furthermore, the removal or suppression of content that documents the realities of the war can obstruct the right to access reliable information, crucial for the public's ability to make informed decisions. When platforms prioritise content moderation over the preservation of critical, verified information, they unintentionally contribute to a cycle of misinformation, leaving users vulnerable to false narratives. In times of war, where the stakes are high, such censorship can undermine public trust, destabilise social cohesion, and silence voices that are vital for healing, accountability, and progress.</p>
<b>Risk-specific recommendations:</b>	<ul style="list-style-type: none"> <li>a) Contextual analysis should be integrated into moderation at all levels, ensuring collaboration with linguists and experts on sociocultural peculiarities and regional specifics. The analysis should be gender-sensitive and place emphasis on context-specific discriminatory discourses and connected terminologies.</li> <li>b) Improve education and training of moderators, focusing on deepening their understanding of cultural, historical, gender and political aspects.</li> <li>c) Ensure consistency between different aspects of moderation and prohibit algorithms from making decisions without real contextual analysis.</li> </ul>

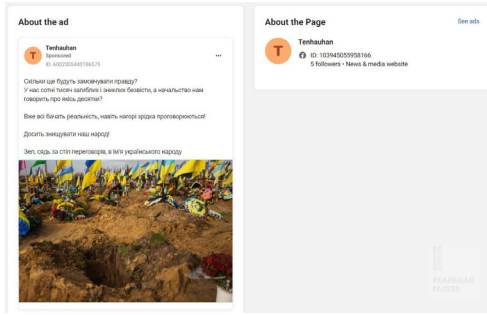
## Case 6: Lack of ability to reach users in temporarily occupied territories of Ukraine.

<b>Date / period:</b>	2022
<b>Platform:</b>	Facebook
<b>Description:</b>	Ukrainian media faced restrictions on content targeting temporarily occupied territories (TOT). This excluded the possibility of targeting Crimea and Sevastopol due to legal restrictions related to sanctions.
<b>Platform response:</b>	Meta explained the restrictions as being necessary to comply with international sanctions but did not offer alternatives ways to inform residents of occupied regions.
<b>Source/ further information:</b>	<p>In an article published in July 2022, the online newspaper <a href="#">Ukrainian Truth</a> documented how Meta's ad restrictions in occupied Ukrainian territories hinder Ukraine's ability to counter Russian propaganda and provide truthful information to civilians there.</p> <p>The screenshot below shows the lack of possibility to target users in the Donetsk region.</p> 




<b>Case relevance:</b>	<p>Residents of temporarily occupied territories found themselves in an information vacuum, which increased their vulnerability to Russian disinformation and complicated their access to reliable information about events in Ukraine. In these conditions, disinformation can easily spread, manipulating public opinion and causing confusion about the situation on the ground.</p> <p>Collaborating with media, civil society, authorities or other relevant stakeholders in an attempt to provide opportunities to reach and target users in regions vulnerable to disinformation and with limited access to information can be crucial to the safety of local residents. Reaching out to users in TOT is a vital part of upholding their right to access truthful, verified information, allowing them to stay connected with the outside world and make informed decisions. Such efforts also help prevent the spread of harmful propaganda and empower individuals to resist manipulation, ultimately strengthening their ability to protect their rights in an unstable environment.</p>
<b>Risk-specific recommendations:</b>	<p>a) To address the complexity, collaboration with local stakeholders, including representatives of TOT, should be established to discuss and find solutions for maintaining the presence of reliable information for residents in temporarily occupied territories. Solutions should be sensitive to the needs of women, minorities, and the marginalised and vulnerable groups.</p>

### Case 7: Abuse of commercial tools for political and military purposes that violate community standards.

<b>Date / period:</b>	2022
<b>Platform:</b>	Facebook
<b>Description:</b>	In 2022, organised campaigns were detected on Facebook with fake pages that used paid advertising to spread identical texts and different visualisations containing political disinformation narratives.
<b>Platform response:</b>	Facebook identified and removed networks of fake accounts and pages that violated the platform's policy on authentic behaviour.
<b>Source/ further information:</b>	<p>The independent investigative media outlet <a href="#">Real Newspaper</a>, in an article published in March 2023, detailed how Russia uses Facebook advertising to spread demoralising and false information to Ukrainian audiences through short-lived, anonymous pages. <a href="#">Page snapshot</a>.</p> <p>The screenshot below shows one example of such pages. It claims that Ukrainian authorities are keeping secret the numbers of Ukrainians who have lost their lives in the war.</p> 
<b>Case relevance:</b>	<p>Disinformation campaigns aiming to undermine morale and discredit authorities can mislead the public, intensify fears and distrust of official information sources, put unnecessary strain on authorities and democratic institutions and thereby contribute to societal destabilisation, which is particularly critical during emergency, crisis and armed conflict. With tools like paid advertisement, and targeted and automatic content promotion, disinformation campaigns have the potential to significantly scale the efficiency and reach of harmful content and campaigns.</p> <p>Sponsored posts or advertising campaigns, for example, promoting narratives about alleged mass losses, aim to undermine morale and demoralise citizens. Such campaigns not only destabilise society but also threaten information security by spreading emotional distress and falsehoods. These actions violate the public's right to truth, endanger public safety, and undermine the ability of individuals to make informed decisions in critical situations.</p>
<b>Risk-specific recommendations:</b>	<p>a) Transparency efforts should be strengthened to publicly provide information about advertising campaigns and buyers of ads.</p> <p>b) Investigations should be performed to ensure that monetisation programmes do not channel revenues to actors associated with sanctioned entities or foreign actors systematically producing and/or distributing disinformation and hate speech narratives.</p> <p>c) An independent audit programme should be developed for advertising campaigns that will include the participation of external experts and civil society organisations.</p>

## Case 8: Gender-based disinformation and hate speech.

<b>Date / period:</b>	December 2024
<b>Platform:</b>	Facebook
<b>Description:</b>	<p>On 19 December 2024, Ukraine experienced a large-scale cyberattack on its state registries, leading to their temporary unavailability. Subsequently, a wave of negative comments and hate speech emerged on social media, particularly on Facebook, targeting the co-founder of the Foundation for Women's Leadership and Strategic Initiatives. The attacks intensified after her participation in an event dedicated to women's role in cybersecurity, where she publicly expressed her position on strengthening gender balance in the technology industry.</p> <p>The attacks had a pronounced gender-based nature. Comments revolved around hostile and gender-stereotypical claims about the lack of competency based on gender. Vulgar and demeaning statements unrelated to her professional activities were aimed at discrediting her as an expert.</p>
<b>The platform's response:</b>	No effective solutions were implemented to address the situation. Complaints from the woman in question and others did not lead to any removal of content.
<b>Source/ further information:</b>	<p>A Facebook post from December 2024 sparked gender-based attacks and hate speech after the co-founder of the Foundation for Women's Leadership and Strategic Initiatives participated in an event highlighting women's roles in cybersecurity. Below are shown examples of posts and more can be found at the original <a href="#">Facebook post</a>.</p> <div>  <p><b>Roman Kurbatov</b> 20 грудня 2024 р. · 🌐</p> <p><a href="#">Mike Sydorenko</a>, посмотри какое вкусное. Если у нас вся кибербезопасность только этим и занята, я вообще не удивлён тому, что мы остались без реестров.</p> <p>“(…) look how delicious. If all our cybersecurity is only occupied with this, I am not surprised at all that we are left without registries.”</p> </div> <div>  <p><b>Roman Budzyk</b> Усі порти у мадам будуть славно відюзані...без термопасті і охолодження</p> <p>9 тиж. Подобається Відповісти</p> <p>“All ports of Madam will be gloriously used.. without thermal paste and cooling”</p> </div> <div>  <p><b>Barsuk Mihalych Barsuk</b> прикольная девушка. то есть, мозгов нету совсем. лидер фейри или квадроберов. но, реально, прикольная. она же ни хрена в словах, которые пишет, не понимает. Кто бы оценил смелость. А не только глупость.. Хотя, потому и смелая, что глупышка.. ик</p> <p>9 тиж. Подобається Відповісти</p> <p>“a cool girl, that is, no brains at all. the leader of the fairies or quadrobbers. but, really, she is cool. she does not understand a damn thing in the words she writes. Who would appreciate courage. And not just stupidity.. Although, that's why she is brave, because she is a fool.., ik”</p> </div>
<b>Case relevance:</b>	The lack of proper moderation and response tools risks exacerbating discrimination and jeopardising safety. Protecting women, minorities and marginalised groups from hate speech and disinformation is essential for upholding human rights, preventing harm and violence, and fostering social cohesion. These groups are often disproportionately targeted, and unchecked hate speech can escalate into real-world harm. Ensuring their safety online promotes equality, freedom of expression, and a healthy digital environment, where all individuals can engage without fear. It also helps maintain public trust, encourages participation, and strengthens the integrity of digital platforms as spaces for meaningful discourse.
<b>Risk-specific recommendations:</b>	<ol style="list-style-type: none"> <li>Collaborate with local, regional and global experts to inform and improve moderation policies and practices to protect women, minorities and marginalised groups.</li> <li>Provide support systems for women, minorities and marginalised groups facing gender-based disinformation and hate speech, such as rapid reporting mechanisms, access to mental health resources, and options to secure their accounts.</li> <li>Additionally, work with NGOs and local women's organisations to support and protect women in the public eye from potential offline harm connected to online abuse.</li> </ol>

## Research

There are two risks that have not been exemplified by concrete cases but are widely recognised as a threat to access to information and are equally relevant to address. They are also addressed as two of the 10 key risks identified by a local group of experts for the “Guide for Risk Management in the context of emergencies, armed conflicts and crises”. The following independent research provides insights into these challenges.

### **Ineffective tools for searching reliable war-related information on platforms.**

The independent Ukrainian media outlet [Texty.org.ua](https://texty.org.ua) has documented that during the ongoing war, Ukrainians have had ineffective tools for searching reliable war-related information on platforms.

In the summer of 2023, 205 Ukrainians volunteered to let the media outlet monitor the recommendations they were provided by YouTube. The outlet concluded that there were several issues related to Ukrainian users being recommended Russian disinformation and hate speech while using the platform.

One main problem is how YouTube’s algorithm promotes content based on language. One user, “Vatnik”, who watched a recommended video by a pro-Russian blogger was subsequently recommended more videos and channels promoting disinformation about the war in Ukraine. The recommendations included interviews and reels with the American pseudo-expert with pro-Russian views, Scott Ritter, who had had his two YouTube accounts closed by the platform for spreading false information. Thus the platform promoted persons and organisations which it has otherwise attempted to ban. Furthermore, another problem detected was that pro-Russian disinformation was often labelled in the wrong category, so users who were searching for music, for example, would be recommended political content and disinformation about the war.

All these issues related to YouTube’s recommendations make it difficult for Ukrainians, who are navigating an information war, to search for and easily find reliable information relevant to the war. It conflicts with their right to access to information during a vulnerable time.

### **Risk-specific recommendations:**

- a) Efforts to collaborate with local media and information literacy organisations and institutions should be furthered, including organisations that focus on a gender-sensitive approach and are familiar with the local gender context.
- b) Labelling should be introduced for accounts that have signs of inauthentic coordinated behaviour. It can be based on several indicators, for example, the use of bots for interaction with content, abnormal activity spikes, geographical location mismatch, name changes as well as analysis of account behaviour, including "renting out" the account for manipulative actions.
- c) Financial and human resources should be enhanced for efficient and qualified handling of appeal cases from users who have wrongly been labelled as an unreliable source. Local stakeholders should be involved in evaluating key cases.

### **Users are recommended harmful content, including disinformation and hate speech.**

According to a [study](#) from November 2024 by the [Center for Strategic Communications and Information Security](#), TikTok is actively used to spread disinformation among the Ukrainian audience. The platform's algorithms promote disinformation with pro-Russian narratives: for example, content discrediting the Ukrainian army. Videos containing false information about mobilisation, false criticism of Ukrainian authorities, and denial of war crimes receive a high number of views through automatic recommendation mechanisms.

An example of this is the anti-Ukraine and pro-Russian social media influencer Anton Gura, who has 48,200 followers. He is documented to have spread disinformation about several topics, such as the introduction of the e-hryvnia and power outages in Ukraine allegedly caused by conflict between the government and national infrastructure companies rather than Russian attacks. His biggest videos on TikTok are his anti-mobilisation videos that get views ranging from tens of thousands up to 1.5 million.

Gura's videos are part of a larger trend: from 16 May–16 June 2024, about 13,000 videos with the hashtag #TCC were published and gained more than 470 million views altogether. TCC stands for Territorial Center of Recruitment and Social Support and has been the focus point of numerous Russian disinformation campaigns attempting to disrupt mobilisation for the Ukrainian military. In July, #TCC became the most popular hashtag in the Ukrainian segment of TikTok along with other anti-mobilisation hashtags such as #stoptsk, #poprattik, #spolotivtsek, #protesttsk, as well as the anti-state #cenomeyUkraina.

Another example are posts praising Russia's president Vladimir Putin - an individual, who, following an investigation into war crimes, crimes against humanity and genocide has a warrant for his arrest issued by the International Criminal Court (ICC) - as the greatest leader of the world being recommended in the feeds of Ukrainians. Such content can serve to justify criminal actions; it conflicts with the importance of accountability and truth-seeking for victims of human rights violations and fails to protect the dignity of Ukrainians who are under Russian invasion.

### **Risk-specific recommendations:**

- a) Companies should collaborate with local media, civil society groups and other relevant stakeholders to support media literacy capacity building among audiences.
- b) Companies should create communities with local experts who can provide early warnings and respond to surges of harmful content and advertising content as well as provide recommendations for moderation.
- c) Trainings for company moderators should be conducted and should focus on identifying gender-motivated disinformation and gender-based violence.