BULLETIN
OF "CAROL I" NATIONAL DEFENCE UNIVERSITY

https://buletinul.unap.ro/index.php/en/

# Risks, threats, and vulnerabilities related to social media platforms and search engines. Regulations and national legal frameworks

**Dănuț MAFTEI, Ph.D.***
**M.A. Student Lorin Nicolae BOGDAN-DUICĂ****

*National Cyber Security Directorate, Bucharest, Romania
e-mail: dn.maftei@gmail.com
**National Cyber Security Directorate, Bucharest, Romania
e-mail: bogdanduicalorin@yahoo.com

## Abstract

Online social media platforms and search engines are used more and more by violent people, criminal offenders, cybercriminals, and other state or non-state malicious actors, who are involved in activities connected to hybrid threats and foreign interference, causing challenges for children, girls, women, citizens, societies, economies, critical services, democracy, and homeland security.
Social media platforms and search engines could do more to address these issues so as to ensure a free, open, safe, secure, and reliable internet for everybody and to maximize its positive effects. Neglecting the proliferation of illegal activities not only erodes trust in online platforms but also places at risk the security and privacy of its users.
To counter efficiently all the challenges, urgent new regulatory frameworks are needed. The regulations should be applied to social media platforms, search engines, and services that allow users to post content online or to interact with each other.

**Keywords:**
social media platforms; national security; democracy; malicious actors; foreign interference; false information; violence; frauds.

**Article info**
Received: 14 November 2024; Revised: 2 December 2024; Accepted: 6 December 2024; Available online: 17 January 2025

Citation: Maftei, D. and L.N. Bogdan-Duică. 2024. "Risks, threats, and vulnerabilities related to social media platforms and search engines. Regulations and national legal frameworks". *Bulletin of "Carol I" National Defence University*, 13(4): 249-265. https://doi.org/10.53477/2284-9378-24-62

The emergence of social media marked a new era for humankind, facilitating the evolution of humanity. Nowadays, there are several social media platforms (SMPs), search engines, and services that allow users to interact fast with each other or to post content online. This includes a variety of websites, apps, and other services, including social media services, video-sharing platforms, consumer file cloud storage and sharing sites, online forums, online instant messaging services, and dating services. They could be used for watching videos, gaining knowledge, sharing special moments, and reconnecting with friends.

On the other hand, these popular services and social media platforms also have a dark side because they can be a hotbed for scams, fraud, violence, disinformation, and false information. Online platforms and new technologies have made it easier, cheaper, and quicker than ever before for both domestic and foreign malign state or non-state actors to put them into practice. Moreover, anonymity and the lack of control and effective content verification mechanisms facilitate the spread of these messages and the identification of attackers. In parallel, the frequency and complexity of Tactics, Techniques and Procedures (TTPs) used by malicious actors to exploit the weaknesses of platforms and users are unceasingly increasing.

Social media platforms enable the rapid and widespread dissemination of false information and other forms of foreign interference that threaten democratic principles and values. They could be used to plan and display violent acts or spread fake news and harmful messages. Because activities conducted on SMPs can undermine democracy, polarize opinion, incite violence, shake trust in institutions, fuel discrimination and marginalization, and erode social cohesion, the impact on society is profound and complex. For instance, even individuals who do not use social media for violent purposes find themselves caught up in violence due to algorithms that are set to promote this kind of content and encourage acts that lead to real-life violence.
Scams originating from fake ads posted on social media have also increased dramatically. Even legitimate ads are cloned and used for malicious purposes, and for the end users it is hard to know whether the ad is legitimate without clicking on it (Alexander 2024).

The number of social media account takeovers is on the rise and the content is not necessarily checked, the reason why no one can be certain about interacting with someone known. Users should approach all social media interactions, whether it is a tweet, a post, or a direct message, with an appropriate dose of skepticism, being hard to recognize the signs an app cannot be trusted. For instance, since **online scams** (Stathis 2024a) and **Facebook Marketplace scams** (Alexander 2024) are so prevalent, it would be wise for users to learn how to identify a scammer (Stathis 2024b).

Malicious actors, especially the state ones, exploit social media platforms as means for conducting hybrid warfare operations. Researchers have especially noted the

evolution of Russian information warfare doctrine, along with its "deep roots in long-standing Soviet practice" (Giles 2016; Snegovaya 2015). The recent Russian military thinking emphasizes hybrid warfare as a new persistent reality, with the "information sphere" and "information warfare" as a critical battlespace.

So-called "trolls" and "bots" seem to play a key role in spreading **fake news** and **disinformation** through social media platforms. Professional trolls manage human-run accounts to provoke or spread disinformation and fake news on social media. On the other hand, bots are involved in managing automated accounts that combine human-generated content with computerized posting. To achieve their objectives, large networks of false accounts are created and used, which is why they play a significant role in promoting fake news.

Methods used by the social media companies to identify automated accounts and coordinated fake news campaigns conducted by state actors are different and the results are also diverse. Although platforms have implemented some content moderation measures, they are often insufficient and slow to respond the requests to remove harmful content. This is partly due to both the large volume of user-generated content and a lack of enough incentives to act quickly and efficiently. Despite its immense resources and technological prowess, *Meta for Business* has been criticized for its inadequate response to the proliferation of phishing scam pages on Facebook. The company's algorithms and content moderation mechanisms are often found lacking in means of identifying and removing these deceptive pages promptly, repeatedly replying to the request of the users that the "*content doesn't go against the community standards*", or that "*it is safe*". This leaves millions of users exposed to the risks associated with phishing scams, and it raises significant questions about Meta's commitment to user safety (Qureshi 2023).

## Negative effects of malicious messages posted on social media platforms

The current scientific research has shown that malicious messages of various types that are posted on social media platforms have a negative impact both on individual users and on national security, democracy, and societal stability.
Foreign actors intend to create the conditions for manipulation or other interference by eroding public trust, destabilizing political systems, undermining democratic norms, and weakening the resilience of democratic states. In the long term, this can damage the ability of democracies to withstand external threats or maintain effective governance.

One of the most damaging effects of Foreign Information Manipulation and Interference (FIMI) is the erosion of public trust in democratic institutions. Misinformation, fake news, and hate speech, including targeting ethnic, religious,

and sexual minorities, amplify social divisions, lead to increased discrimination and violence against them, and fuel political and cultural polarization.
At the same time, trust in institutions and traditional media is eroded, leading to increased skepticism and difficulties in distinguishing between real and false information.

By exacerbating existing divisions in society, FIMI campaigns amplify the polarization of political discourse, making it more difficult for democratic societies to engage in constructive debate or identify common ground on the critical issues facing them. This polarization weakens the ability of democratic institutions to function effectively, turning the legislative process into partisan gridlock and political extremism.

SMPs, disinformation, hate speech, fake news, and cyber-attacks used to manipulate voters' opinions, as well as to increase social tensions and the level of violence before, during, and after elections, can influence the outcome of elections. Such activities can have far-reaching consequences as they can lead to the election of candidates who are more favorable to foreign interests or, conversely, harm the prospects of candidates perceived as hostile to such interests.
Women and teenage girls fall prey to various forms of online sexual violence (cyberbullying, rape videos, threats, and distribution of sexual images without consent). These forms of violence can become real and interfere with women's ability to feel safe at work or in public.

On the other hand, the popularity and ease of use of SMPs have made it easier for extremists to access like-minded people, create terrorist networks, recruit new members, spread extremist ideologies, and incite violence. SMPs algorithms can amplify extremist content, exposing users to dangerous messages that contribute to their radicalization. Foreign interference through health-related misinformation (e.g., on vaccines, pandemics, etc.) harms public health, increasing the risk of illness and premature death.
At the same time, misinformation and fake news can negatively affect national and international economies by manipulating financial markets, causing financial damage, undermining business confidence, and spreading panic.
Particular challenges are also posed by human trafficking for labor and sex trafficking, with children and even adolescents and young adults being the most common victims.

## Tactics, Techniques, and Procedures used online by malicious actors for conducting fraudulent activities

Day after day both old and new TTPs can be seen that are used by scammers to trick people. With more and more online frauds being carried out every day, each

new fraud is even more complex, cleverer, and less detectable than the last one (Stathis 2024a). Currently, SMPs users are victims of several types of threats, which are summarized below.

**Social media phishing** (Adrien 2023) means attacks conducted online. Their purpose is to steal personal data or gain control of social media accounts. Phishing is a form of cybercrime where malicious actors impersonate trustworthy entities often to lure users through fake promotions, fraudulent contests, or fabricated news stories, or to deceive people into clicking on malicious links or revealing sensitive information such as personal details, login credentials, credit card information, financial data, etc. Phishing activities are connected to cybercrime, and they currently represent one of the most common forms of social engineering, with more than three billion spam emails sent every day.

According to statistics, millions of Facebook business accounts worldwide are being targeted with phishing messages, with a success rate of nearly one in seventy victims infected (Petkauskas 2023). Fraudsters usually impersonate SMPs in phishing attacks designed to sneak malicious software (spyware or ransomware) onto personal computers, and steal login information and potentially other personal data (Rosenkrantz 2024).

Phishing remains popular, but we could notice currently new phishing techniques like *spear phishing, whaling, business email compromise, smishing, https phishing, clone phishing, pop-up phishing, angler phishing, evil twin phishing, search engine phishing, watering hole phishing, vishing,* etc. (Chin 2024). Moreover, cybercriminals use generative artificial intelligence tools to write their emails, which significantly improves their phishing success rates.
Hackers use a massive network of fake and compromised accounts to send out millions of Messenger platform phishing messages to target Facebook business accounts with password-stealing malware (Toulas 2023b). According to reports (Zaytsev 2023), researchers warn that roughly one out of seventy targeted accounts is ultimately compromised, translating to massive financial losses.

**Fraudulent applications** (Budgar 2024) can be advertisements for apps or features on SMPs that claim to allow users to check who has viewed their profile.
In the case of the **Facebook Marketplace scams**, it can be seen that a huge number of users buy and sell goods every day, but also that scammers use this online shopping platform to scam people and steal their money (Alexander 2024). Scammers may ask users to pay or discuss additional details by using third-party communication channels, while others might list fake rentals, gifts, or various products.
In **bank fraud**, many scammers offer **fake gifts** to get users to divulge various personal information (credit card, social security numbers, etc.) or to access links where they can download viruses on their personal computers (Bradford 2024).

In **spoofing attacks**, hackers can illegally access a person's account and then send fake messages or posts to their friends asking for money or gifts (Alexander 2023). The messages are designed to excite or panic the user and then get them to provide money without properly analyzing the situation. In addition to using a friend's profile to conduct a spoofing attack, scammers might impersonate famous people or organizations.

**Sextortion** is a social engineering scam where a victim (usually male) is befriended by a female scammer, convinced to send sexually explicit images or videos of themselves to the fake persona, who then threatens to release live the compromising material if the victim refuses to pay up (Schappert 2024).

Attackers may also use **"Secret Santa" schemes** where people send a $10 gift to one person and then receive one from three others. But there is no guarantee that the victim will get their money back in these Facebook scams, because if no one follows through on sending the gift, they may get nothing in return. Malicious actors could use the victim's home address to carry out *doxxing attacks*[1] (Alexander 2022), and sharing other personal information could reveal the answers to password security questions, leaving personal accounts vulnerable to hackers.

[1] *Doxxing* or *doxing* is the act of publicly providing personally identifiable information about individuals or organizations, usually online and without their consent, as a form of punishment or revenge.

**Misinformation** refers to **false information**, misleading, or taken out of context, disseminated by *a p*erson who *believes it is true*, without intention to cause harm. Misinformation has the power of "social proof" in persuading individuals to accept false information. People could accept faster news stories as true when they are disseminated by friends, acquaintances, and supposedly credible sources, and when these stories are more popular overall (Hindman 2018).

**Disinformation** (PakVoices 2023) refers to **false information** (or manipulated narrative or facts, propaganda), and the *propagator knows it is false*. It is a deliberate, intentional lie, intended to manipulate, cause damage, and guide people, organizations, and countries in the wrong direction, generating mistrust in the democratic state institutions, either for the purposes of causing harm, or for political, personal, or financial gain.
Disinformation has multiple stakeholders involved; it is coordinated and hard to track. It may include doctored videos, fake news articles, and artificially amplified social media posts. It often contains slander or hate speech against certain groups of people and is often polarizing, inciting anger and other strong emotions and it can lead people to promote extreme views, and conspiracy theories, without room for compromise.
New emerging technologies are increasingly used to discredit factual information. Artificial intelligence (AI) and generative AI may be used to spread false and misleading information, such as "*deepfake*".

**Malinformation** refers to *reality-based information* that is used to harm individuals, social groups, organizations, or nations (ITU 2021). Malinformation involves real, not false, facts. Personal data and leaked emails revealed through *doxxing* are examples of malinformation. Harassment, hate speech, and revenge porn also fall into this category.

**Fake news** is *purposefully crafted*, sensational, emotionally charged, misleading, or totally fabricated information that mimics the form of mainstream news (Saint Francis University 2023). They are used for the online distribution of false information disguised as legitimate news stories. Motivations behind fake news could be personal (to harm an individual/business reputation), financial (to attract internet traffic and/or advertising income), or political (to influence the public's viewpoint/ideology).

***

Of course, there are plenty of other variations of challenges that people can face on the SMPs, such as *malware attacks, spam messages, cloned accounts, fake medical fundraising, clickbait scams, fake coupon code scams, Facebook quiz scams, romance scams, job scams, fake fundraising, cyber stalking, internet banging, child sexual abuse, control or coercive behavior, extreme sexual violence, extreme pornography, sale of illegal drugs or weapons, sexual exploitation, fraud, racially or religiously aggravated public order offenses, illegal immigration and human trafficking, promoting or facilitating suicide, abuse of intimate images (revenge porn), terrorism*, etc.

**It needs to be clearly understood by decision-makers** that all these types of TTPs represent an enormous number of possible ways of action that can be successfully used by various malicious actors to conduct complex online attacks with serious results. Following detailed scans of victims to identify their specific vulnerabilities, the attacks will then be organized, tailored, and customized exactly to the specifics of each target, combined with other state-of-the-art methods and technologies, so that the chances of success are maximized. As such, under these conditions, states need to adapt quickly by amending the legal framework and developing effective working strategies to counter such complex challenges.

## Specific measures taken by national authorities for combating the illegal activities conducted by using social media platforms

It could be noticed that the EU and different countries around the world have been paying attention for years to the malign activities conducted on SMPs and to their impact on national security, democracy, state institutions, critical infrastructure, society, businesses, and citizens. The current research has highlighted several measures taken by national authorities against challenges posed by malicious actors (state and non-state) using SMPs, as follows.

***The TikTok platform:***

Since 2020, the TikTok Platform has been blocked/restricted in countries such as Afghanistan, Armenia, Azerbaijan, Bangladesh, India, Iran, USA, the reasons behind these decisions being related to national security, high levels of terrorism, border conflicts, etc. (Gordon 2024).

Under the Digital Services Act[2], the European Commission opened proceedings against TikTok over the launch of *TikTok Lite* in France and Spain (European Commission 2024).

[2] The Digital Services Act Regulation mandates that digital platforms take greater responsibility for the content shared on their platforms. This legislation seeks to limit the spread of harmful disinformation while ensuring that freedom of speech is respected.

In 2023, TikTok was banned on devices owned by state institutions in Austria, Belgium, Canada, Estonia, France, USA, due to security and privacy risks, as well as alleged links between the Chinese Communist Party and the company, with TikTok accused of collecting and sharing personal data with Chinese intelligence services (Lakshmanan 2024).

In May 2023, in Romania, the National Cyber Security Directorate – DNSC (a specialized body of the central public administration, under the authority of the Government, responsible for ensuring the cyber security of national civilian cyberspace), issued a recommendation to national state institutions and public bodies not to download, install and use TikTok on their networks and information systems (DNSC 2023b).

Taiwan had banned TikTok from government devices in December 2022. The reason was connected to concerns of it being used by China to carry on ”cognitive warfare” against Taiwan.

The technical reports on TikTok show the presence of *a lot of cybersecurity risks and vulnerabilities related to installing and using this application* (**collecting personal data, used devices, operating system, IP, SSID Wi-Fi, Serial number, SIM ID, IMEI, SMS reading, MAC Address, GPS location, user accounts, clipboard access, history, useless Do Not Track setting, services/applications used, user personal profiling, sharing collected data to other ”partners”, remote control, etc**.) (Baiaș 2023).

Meanwhile, the Chinese legal framework, which obliges citizens and entities to cooperate with intelligence services and state institutions to provide data and information for “national purposes”, was taken into account (***The State Security Law*, 2015; *The Cybersecurity Law*, 2016; *The Law on State Intelligence Activities*, 2017; *The Law on State Counterintelligence Activities*, 2023**).

***The Facebook platform:***

Since 2015, Facebook has been blocked in Ethiopia, Bangladesh, Myanmar, and Sri Lanka to prevent the spread of disinformation and hate speech, control the flow of information, and suppress dissent, on national security grounds or because of content deemed offensive to Islam.

Facebook, on the other hand, has been subject to restrictions and censorship in China, Iran, and North Korea, where access to the platform is either completely blocked or severely restricted.

***Instagram***:
Instagram has been blocked in China since 2014 as part of the Chinese government's efforts to control the flow of information and limit access to Western social media platforms. Meanwhile, Instagram was blocked intermittently in Iran during political unrest and protests to prevent and stop the spreading of information and coordination of demonstrations.
Turkey temporarily blocked access to Instagram and other SMPs after an attempted coup to prevent the spread of misinformation and panic (2016).

In 2020, India banned Instagram and around sixty other Chinese apps, citing national security and data privacy concerns (2020). In the same year, the Russian Federation blocked Instagram as a response to Meta's decision to allow users in certain countries to post calls for violence against Russian soldiers in the context of the war in Ukraine.
Instagram has also been subject to restrictions and censorship in North Korea and Turkmenistan, where internet access is strongly controlled by the national government.

## The current legal framework issued by EU/EU member states and non-EU countries for regulating social media platforms

The European Union and several countries around the world have been paying attention to legal and regulatory frameworks to make the use of internet services safer for citizens, organizations, and businesses, but also to make social platforms more accountable. These laws impose obligations regarding transparency, content moderation, and response to requests from authorities. Moreover, authorities responsible for the activities of SMPs have been established.

In contrast, in other countries, the legislation needed to regulate social media platforms is inadequate or non-existent. This leaves authorities without effective tools to compel platforms to take responsibility for hosted content and promptly respond to requests to remove harmful content.

In the **European Union**, the European External Action Service has been working since 2015 on tackling FIMI, including disinformation, and on strengthening its strategic communications in the Eastern Partnership, the Southern Neighbourhood, and the Western Balkans (EEAS 2024). To this end, the General Data Protection Regulation

(GDPR) - April 27, 2016 (EUR-Lex 2016), the Digital Services Act (DSA) - December 15, 2020 (EUR-Lex 2022b)[3], and the Digital Markets Act (DMA) - December 15, 2020 (EUR-Lex 2022a)[4] have been developed.

[3] Regulation (EU) 2022/2065 on a Single Market for Digital Services.
[4] Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector.

**Germany** has also shown interest in adapting legislation to current challenges. To this end, the **Network Enforcement Act (NetzDG)** was adopted in 2017 (bundesjustizamt.de 2018). Germany's NetzDG law is **one of the strictest regulations in Europe for combating online hate speech and disinformation**. The law obliges SMPs that have more than two million users in Germany to remove unlawful content within 24 hours or face fines of up to €50,000,000. While not solely focused on foreign disinformation, NetzDG plays a key role in preventing the spread of foreign-originated manipulative content

Meanwhile, Germany established the *National Cyber Defense Center*. This joint institution includes representatives from federal agencies, including the *Federal Office for Information Security* (BSI), the *Federal Intelligence Service* (BND), and the *Federal Office for the Protection of the Constitution* (BfV) – Germany's domestic intelligence agency. The Center coordinates Germany's response to cyber threats, which include FIMI and the use of cyber tools to spread disinformation.
The BfV has developed specialized programs to monitor FIMI in elections, particularly focusing on Russian and Chinese disinformation campaigns. Ahead of the 2021 federal elections, the BfV issued warnings and enhanced its monitoring of SMPs, and foreign-funded groups involved in spreading disinformation.

**France** adopted the 2018 *Law against the manipulation of information (Loi contre la manipulation de l'information)*, as a response to increasing concerns about FIMI in elections. Known as the "*Fake News Law*", it enables judges to act swiftly during elections by removing/blocking disinformation from media sources if it can be proven that they deliberately spread misleading information aimed at manipulating the outcome of an election. The law also requires SMPs to disclose their sponsors during election campaigns to avoid foreign-financed manipulation.

At the same time, *the Higher Audiovisual Council (Le Conseil Supérieur de l'Audiovisuel – CSA)*, France's media regulatory body has been granted enhanced powers to oversee media platforms and content dissemination. During election periods, the CSA can act against platforms that allow the spread of disinformation or manipulation originating from foreign actors. Furthermore, the CSA can impose sanctions on outlets that fail to meet transparency standards regarding political advertising.

For the regulation of social media platforms, the **US** makes use of Section 230 of the *US Communications Decency Act*, February 21, issued in 1996 (LLI 1996).
Starting with 2021, in **Australia** the *News Media Bargaining Code* is functioning.

In the **United Kingdom of Great Britain and Northern Ireland**, the *Online Safety Act 2023* was enacted (the Act) (GOV.UK 2023a). The Act is a set of strong regulations for protecting children and adults online. It contains new strict regulations for online SMPs and search engines, including obligations to protect users from harmful content, to quickly remove illegal content, to implement needed systems and processes for reducing risks related to their services when used for illegal/malicious activities, to take down illegal content as well criminal offenses. The law also contains provisions relating to *Ofcom* (the *Independent Regulator of Online Safety*), an entity involved in setting out the steps that providers can take to fulfill their safety duties in codes of practice, and has a broad range of powers to assess and enforce providers' compliance with the framework.

The Act's duties apply to search services/engines and services that are used to *allow users to post content online or to interact with each other*. This includes a range of websites, online instant messaging services, apps and other services, social media services, consumer file cloud storage and sharing sites, online forums, video-sharing platforms, and dating services. The Act applies to services linked to the UK, even if the companies providing them are outside the country (GOV.UK 2023b). The criminal offences introduced by the Act apply directly to the individuals sending them and cover: encouraging or assisting serious self-harm, cyberflashing, threatening communications, sending false information intended to cause non-trivial harm, intimate image abuse, epilepsy trolling.

The specific illegal content and activities, that platforms need to protect users from, are related to child sexual abuse, extreme sexual violence, controlling or coercive behavior, extreme pornography, fraud, inciting violence, racially, and religiously aggravated public order offenses, illegal immigration, and people smuggling, promoting or facilitating suicide, selling illegal drugs or weapons, intimate image abuse (revenge porn), sexual exploitation, terrorism. The Act also requires the platforms to rapidly remove illegal suicide and self-harm content and proactively protect users from content that is illegal under the *Suicide Act* from 1961.

The UK's *Counter Disinformation Unit* (CDU) was set up in 2019, being focused on monitoring the online content that poses risks to public health, public safety, and national security and responding to risks of misinformation, including that on Covid-19. CDU is involved in analyzing disinformation attempts and could work with social media companies to encourage them to promote authoritative sources of information. Currently, it is focused on disinformation related to Russia's illegal invasion of Ukraine and has already countered Russian disinformation about Ukraine. CDU has been talked about more than two hundred times in the British Parliament.

**Canada** operates the *Online Harms Bill*, which aims to tackle harmful online content, including hate speech, misinformation, and child sexual abuse.

**Singapore** enacted the 2019 Law for *Protection from Online Falsehoods and Manipulation Act* (POFMA) (Singapore.gov 2019). It allows the government to order the correction or removal of false/harmful information from SMPs.

**Brazil** is also paying attention to online activity, with the 2014 *Brazilian Civil Rights Framework for the Internet Law* (Secretaria-Geral 2014) that establishes principles for using the Internet in Brazil, including net neutrality and the protection of personal data.

In **India**, the Information Technology Rules (*Intermediary Guidelines and Digital Media Ethics Code*) or the "*IT Rules*" (Indian.gov 2021) came into effect in 2021 and laid down some specific compliance requirements for social media intermediaries. The *IT Rules* were introduced to check the spread of fake news, hate speech, and online harassment, some of the significant aspects being as follows:

- The SMPs/other intermediaries have to observe due diligence by making reasonable efforts to cause their users not to host, display, upload, modify, publish, transmit, store, update, or share any information that (1) is harmful to children (2) infringes the trademark, copyright, patent or other proprietary rights (3) is defamatory, obscene, invasive of the privacy of another person, is racially or ethnically objectionable (4) impersonates another person (5) violates any law.
- The rules provide an effective redressal mechanism by which users/victims may submit a complaint against IT Rules violations. The Grievance Officer must act in a time-bound manner after receiving a complaint in a request for the removal of information or communication link.
- It is mandatory for all significant SMPs to appoint a *Chief Compliance Officer* and a *Nodal Officer* who would be available 24*7 for coordination with law enforcement agencies.

In **Romania**, in May 2023, DNSC issued a recommendation to national state institutions and public bodies not to download, install, and use TikTok on their networks and information systems. At the same time, Romanian authorities are considering new legal provisions aimed at stricter regulations for social networks, creating a safer and more responsible online environment, designating national contact points/representatives for social networks in Romania, and introducing sanctions for non-compliance with content moderation obligations.

## Conclusion

Online platforms and search engines allow users to develop global networks and are currently the most popular medium among content creators. The concept behind them seems innocuous, but the ease of access and the opportunities they offer also

involve some risks. Abuse of intellectual property, theft of personal and banking data, misinformation, spreading fake news, obscene content, violence, or hate speech are some of the challenges.

Both malicious activities conducted on SMPs by state and non-state actors and other forms of foreign interference constitute a threat to democratic principles and values, having a negative impact on national security, democracy, state institutions, critical infrastructure, society, business, and citizens. Some of those most exposed to harmful and inappropriate online content are children, women, girls, but also the elderly.

This scientific research attests that both ordinary users and national authorities face problems related to the lack of a legal regulatory framework, formal procedures, or the possibility to directly contact representatives of social media platforms when needed to take action to block/remove/modify such illegal activities or inappropriate messages in a timely manner. The study also notes several complaints of lack of adequate response from MSPs to user reports and requests to block/remove attack vectors.

On the other hand, SMPs face ongoing challenges in moderating the illegal content published online. Some of them have implemented various measures to address these issues (content moderation; increasing transparency around content moderation; improving algorithms to automatically detect harmful content; classifying content; acting on the resulting classifications; and working with fact-checkers), but they are often insufficient and slow to respond to requests to remove harmful content. This is partly due to the large volume of user-generated content, but also due to a lack of sufficient incentives and penalties to act quickly and efficiently.

The time has come for SMPs to recognize their responsibility, invest in robust security measures, proactively tackle this, and prioritize the safety of their users in the digital age. Legal obligations should be brought to the attention of all audio-visual broadcasters and social media platforms to provide the public with unbiased and objective information, presenting facts and events accurately, while respecting the freedom of expression.

After this present study, a main conclusion could be drawn: **countries need effective regulatory frameworks and policies for making the use of Internet services safer**. They should be applied to **social media platforms, search engines, and services that allow users to post content online or to interact with each other**: a range of websites, online instant messaging, online forums, services apps, and other services, including social media services, consumer file cloud storage and sharing sites, video sharing platforms, dating services, etc. The legislation should be balanced, protect freedom of expression, but also ensure that online platforms take responsibility for the content they host and contribute to a safer and healthier online environment, to protect users from harmful content, quickly remove illegal content, implement

systems and processes necessary to mitigate the risks of the services offered when used for malicious activities.
In light of the current context and international experience, countries around the world could consider taking legislative action in the following areas to better regulate social media platforms, search engines, and services that allow users to post social content online, and to protect them.

Looking at the current framework, **transparency and accountability** of online platforms are essential elements. They should appoint a national representative in the countries where they operate, responsible for communicating with the authorities and ensuring compliance with local legislation. Users also need simple and accessible mechanisms to report harmful content and/or challenge moderation decisions. At the same time, online platforms should regularly publish detailed reports on the measures taken to moderate content, the number of complaints received, and how they have been resolved.

In terms of **content moderation**, social media platforms, search engines, and services that allow users to post content online or interact with each other should be obliged to remove illegal content within a short period of notice. They also need to work better with independent fact-checking organizations and human rights experts to improve content moderation. On the other hand, platforms should be encouraged to use advanced technologies such as artificial intelligence to quickly identify and automatically remove harmful content.

Regarding **user protection**, it is important to implement specific measures to protect children from harmful content, such as age restrictions and parental control tools. SMPs should take effective measures to limit the spread of misinformation, with an emphasis on labeling false or misleading content and promoting credible sources of information. Personal data protection legislation must be strictly enforced, and users must be in control of how their data is collected and used.

To support these measures, **supervisory and regulatory bodies need to be established**. Such bodies are needed to oversee the activity of social media platforms, search engines, and services that allow users to post content online or interact with each other. They should also be empowered to act against companies or platforms that allow FIMI or other illegal activities to take place online, and to impose sanctions for violations of the laws and rules imposed.

In terms of **sanctions**, online platforms that do not comply with legal obligations should be subject to fines, proportionate to the seriousness of the infringement and the company's turnover. In particular cases, the national authorities should have the possibility to temporarily suspend or block the services provided by online platforms and search engines, whenever the situation so requires.

## References

**Adrien, Claudia**. 2023. "Phishing Attacks Target Facebook, Microsoft, Making Them Most Impersonated Brands". https://www.channelfutures.com/security/phishing-attacks-target-facebook-microsoft-making-them-most-impersonated-brands.

**Alexander, Brooke Nelson**. 2022. "What Is Doxxing, and How Does It Set You Up to Be Hacked?" https://www.rd.com/article/what-is-doxxing/.

**—.** 2023. "What Is Spoofing, and How Can You Spot It?". https://www.rd.com/article/spoofing/.

**—.** 2024. "14 Facebook Marketplace Scams to Watch Out For". https://www.rd.com/article/facebook-marketplace-scams/.

**Baiaș, Ionuț**. 2023. „Directoratul Național de Securitate Cibernetică recomandă interzicerea TikTok pe dispozitivele instituțiilor publice". https://hotnews.ro/directoratul-national-de-securitate-cibernetica-recomanda-interzicerea-tiktok-pe-dispozitivele-institutiilor-publice-64785.

**Bradford, Alina**. 2024. "8 Common Bank Scams to Watch Out For". https://www.rd.com/list/bank-scams/.

**Budgar, Laurie**. 2024. "Can You Really See Who Viewed Your Facebook Profile Recently?". https://www.rd.com/article/who-viewed-my-facebook-profile/.

**bundesjustizamt.de**. 2018. "Network Enforcement Act Regulatory Fining Guidelines". https://www.bundesjustizamt.de/SharedDocs/Downloads/DE/NetzDG/Leitlinien_Geldbussen_en.pdf?__blob=publicationFile&v=3.

**Chin, Kyle**. 2024. "19 Most Common Types of Phishing Attacks in 2024". https://www.upguard.com/blog/types-of-phishing-attacks.

**Comisia Europeană**. 2024. "Commission opens proceedings against TikTok under the DSA regarding the launch of TikTok Lite in France and Spain". https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2227.

**DNSC**. 2023a. „ALERTA: Tentative de fraudă promovate prin anunțuri sponsorizate pe rețelele sociale". https://www.dnsc.ro/citeste/alerta-tentative-de-frauda-promovate-prin-anunturi-sponsorizate-social-media.

**—.** 2023b. "Press release." https://dnsc.ro/vezi/document/comunicat-de-presa-dnsc-recomanda-autoritatilor-si-institutiilor-publice-din-romania-interzicerea-descarcarii-instalarii-si-utilizarii-a-aplicatiei-tiktok-pe-dispozitivele-de-serviciu-pdf.

**EEAS**. 2024. "Tackling Disinformation, Foreign Information Manipulation & Interference". https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en.

**EUR-Lex**. 2016. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data". https://eur-lex.europa.eu/eli/reg/2016/679/oj.

**—.** 2022a. "Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector". https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32022R1925.

**—.** 2022b. "Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services". https://eur-lex.europa.eu/eli/reg/2022/2065/oj.

**Giles, Keir**. 2016. "Handbook of Russian Information Warfare". https://www.ndc.nato.int/news/news.php?icode=995.

**Gordon, Anna**. 2024. "Here's All the Countries With TikTok Bans as Platform's Future in U.S. Hangs In Balance". https://time.com/6971009/tiktok-banned-restrictions-worldwide-countries-united-states-law/.

**GOV.UK.** 2023a. "Online Safety Act 2023." https://www.legislation.gov.uk/ukpga/2023/50/enacted.

**—.** 2023b. "What the Online Safety Act does." https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer#:~:text=The%20Online%20Safety%20Act%202023,users'%20safety%20on%20their%20platforms.

**Hindman, Matthew**. 2018. "Disinformation, 'Fake News' and Influence Campaigns on Twitter." https://knightfoundation.org/reports/disinformation-fake-news-and-influence-campaigns-on-twitter/.

**Indian.gov**. 2021. "The Information Technology (Intermediary Guidelines and Digital Media Ethics Code)." https://www.meity.gov.in/writereaddata/files/Information%20Technology%20%28Intermediary%20Guidelines%20and%20Digital%20Media%20Ethics%20Code%29%20Rules%2C%202021%20%28updated%2006.04.2023%29-.pdf.

**ITU**. 2021. "Session 5: Disinformation, misinformation, malinformation and Infodemics: Ways to handle". https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/Pages/Events/2021/ASP%20Regional%20Dialogue%20on%20Digital%20Transformation/Session%20Pages/RD-Session-5.aspx.

**Justia**. 2021. "Gonzalez v. Google, LLC, No. 18-16700 (9th Cir. 2021)." https://law.justia.com/cases/federal/appellate-courts/ca9/18-16700/18-16700-2021-06-22.html.

**Lakshmanan, Lavie**. 2024. "Canada orders TikTok to shut down Canadian operations over security concerns". https://thehackernews.com/2024/11/canada-orders-tiktok-to-shut-down.html?m=1.

**LLI**. 1996. "47 U.S. Code § 230 – Protection for private blocking and screening of offensive material." https://www.law.cornell.edu/uscode/text/47/230.

**PakVoices**. 2023. "Disinformation impacts on digital sphere in Pakistan (May-July 2023)." https://pakvoices.pk/?p=13745.

**Petkauskas, Vilius**. 2023. "Facebook Messenger phishing attack pumps out 100K+ weekly messages". https://cybernews.com/news/facebook-messenger-phishing-attack/.

**Qureshi, Anees**. 2023. "Meta Neglecting the Proliferation of Phishing Scam Pages on Facebook, Leaving Millions of Users Vulnerable". https://www.linkedin.com/pulse/meta-neglecting-proliferation-phishing-scam-pages-facebook-qureshi-dsizf/.

**Rosenkrantz, Holly**. 2024. "What Is Phishing, and How Can You Prevent This Cyberattack?". https://www.rd.com/article/what-is-phishing/.

**Saint Francis University**. 2023. "Misinformation, Disinformation, and Fake News". https://libguides.francis.edu/fake-news.

**Sasnauskas, Mantas**. 2023. "We uncovered a Facebook phishing campaign that tricked nearly 500,000 users in two weeks". https://cybernews.com/security/we-uncovered-a-facebook-phishing-campaign-that-tricked-nearly-500000-users-in-two-weeks/.

**Schappert, Stefanie**. 2024. "Meta deletes 63K sextortion scam accounts from Instagram, Facebook".https://cybernews.com/news/meta-deletes-63k-sextortion-scam-accounts-instagram-facebook/.

**Secretaria-Geral**. 2014. "LEI Nº 12.965, DE 23 DE ABRIL DE 2014." http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm.

**Singapore.gov**. 2019. "Protection from Online Falsehoods and Manipulation Act 2019." https://sso.agc.gov.sg/Act/POFMA2019.

**Snegovaya, Maria**. 2015. "Putin's Information Warfare in Ukraine: Soviet Origins of Russia's Hybrid Warfare, Institute for the Study of War." https://www.understandingwar.org/report/putins-information-warfare-ukraine-soviet-origins-russias-hybrid-warfare.

**Stathis, Jaime**. 2024a. "14 Online Scams You Need to Be Aware Of—and How to Avoid Them". https://www.rd.com/list/how-to-avoid-online-scams/.

—. 2024b. "9 Red Flags You're About to Click on a Fake Social Media Ad". https://www.rd.com/list/fake-ads-on-social-media/.

**Toulas, Bill**. 2023a. "Facebook disrupts new NodeStealer information-stealing malware." https://www.bleepingcomputer.com/news/security/facebook-disrupts-new-nodestealer-information-stealing-malware/.

—. 2023b. "Facebook Messenger phishing wave targets 100K business accounts per week". https://www.bleepingcomputer.com/news/security/facebook-messenger-phishing-wave-targets-100k-business-accounts-per-week/.

**Zaleznik, Daniel**. 2021. "*Facebook and Genocide: How Facebook contributed to genocide in Myanmar and why it will not be held accountable*". https://systemicjustice.org/article/facebook-and-genocide-how-facebook-contributed-to-genocide-in-myanmar-and-why-it-will-not-be-held-accountable/.

**Zaytsev, Oleg**. 2023. "«MrTonyScam» — Botnet of Facebook Users Launch High-Intent Messenger Phishing Attack on Business Accounts". https://labs.guard.io/mrtonyscam-botnet-of-facebook-users-launch-high-intent-messenger-phishing-attack-on-business-3182cfb12f4d.