

Mezei, Kitti¹ - Szentgáli-Tóth, Boldizsár²

Some comments on the legal regulation on misinformation and cyber attacks conducted through online platforms³

¹ Research Fellow, Centre for Social Sciences, Institute for Legal Studies; Assistant Professor, Budapest University of Technology and Economics, Faculty of Economics and Social Sciences, Business Law Department; Researcher, University of Public Service, Eötvös József Research Centre, Cybersecurity Research Institute: szentgali-toth.boldizsar@tk.hu

² Senior Research Fellow, Centre for Social Sciences, Institute for Legal Studies; Researcher, University of Public Service, Eötvös József Research Centre, Information Society Research Institute; Associate Lecturer, Eötvös Loránd University, Faculty of Law: mezei.kitti@tk.hu

³ This paper is funded by the Ministry of Innovation and Technology and the National Research, Development and Innovation Office within the framework of the NKFIH grant No. 128976; a138965 ; and the National Laboratory for Artificial Intelligence and the NKFIH grant No. 2019-2.1.11-TÉT-2020-00243.

Abstract

The motivation behind the malicious use of modern technologies can often be directly aimed at influencing public discourse: generating manipulated positions or exaggerating or even relativising positions that are already present in the debate. For example, posts or comments generated by automated tools ("bots") can give readers a false impression of the positions current in society on a given topic or of their actual weight and support. In addition, the rise of deepfake technology based on artificial intelligence poses a new challenge. In addition to the damage to individuals, deepfake can contribute to distorting democratic decision-making and manipulating the electoral process, eroding public trust and exacerbating societal divisions. It can also raise criminal law issues.

Cybercrime is causing increasing social and economic damage, violating the fundamental rights of individuals and threatening the rule of law and the stability of democratic societies. Online platforms provide a new arena for abuse. Users are often unaware of the dangers of their virtual presence and the consequences of the information they share. Therefore, cyber-attacks against online platforms and their users often aim to obtain personal data or other sensitive information that can be used to commit further crimes (e.g. fraud, extortion, etc.). In addition, social networking sites are potential vehicles for distributing malware and illegal content. All these phenomena are particularly dangerous in an epidemic where almost all of us spend much more time online than before.

This study looks at the main possible ways forward for legislation to address this complex problem. What are the tools to effectively manage the new threats to the free flow of opinions to protect this essential precondition for a pluralist social and political system? And, as a basis for action at the community level, how can we protect users who use internet platforms to inform themselves on issues of public interest from disinformation attacks through the cyberspace?

Effective action against cyber-attacks that adversely affect certain fundamental rights requires a combination of instruments, creating the technological, economic, human and legal conditions for meaningful counter-measures. In legal terms, the guarantees that platform providers must offer each user to prevent cyber attacks and illegal content should be laid down, and legal instruments should be put in place to ensure they are always available. In addition, in the event of misusing any content shared on the platform or of personal data made available to the operator, clear responsibilities should be established, and the extent to which the responsibility for protection lies with the platform operator or the user should be clarified. In our study, we outline regulatory options to address these challenges.

Keywords: cyber attacks; freedom of expression; online platforms; marketplace of ideas; digital services act

1. Introduction

During the recent years, social media has become the main channel for our digital communication and now includes several platforms that are also used by "cybercriminals". They have already appeared on Facebook, Instagram, Twitter, Snapchat, WhatsApp and most recently, Telegram, a messaging app with bot functionality. Among online platforms, some social media platforms have, therefore, undoubtedly become new arenas for cybercrime. The following cases will be presented in detail under this heading: cyber-attacks (hacking and malware), phishing and data leakage, online fraud and extortion, and the dangers inherent in using deepfake technology.

In the second part of the paper, we will look at how to effectively counter the new threats to the free flow of opinion in social media to protect this essential precondition for a pluralist social and political system. Furthermore, how to protect users of these Internet platforms who are informed on issues of public interest from disinformation attacks from cyberspace as a basis for action at the community level.

Many important motivations for cybercrime have been analysed in detail in the relevant literature,⁴ and we will briefly discuss several of them in our paper. However, shaping public opinion, influencing decision making and political gain are motives that have been less discussed so far when researching the background of attacks in the online space. Perhaps this is also why, given the issue's importance, few studies have so far focused on the impact of new technologies on the exercise of freedom of expression and the marketplace of political opinion.⁵ However, there is a growing need for a more in-depth understanding of this context, and it is becoming increasingly clear that the impact of new technologies on public debate may continue to grow in the coming period. This is why we believe it is essential for experts working on current issues of freedom of expression and cybercrime to work together to explore the complexities of this combined area and formulate common proposals on the principles and specific content of the relevant legislation. For this purpose, the relevant case law of the European Court of Human Rights (hereinafter: ECtHR) will be also discussed.

⁴ For more on this, see David S. Wall, *Cybercrime, Media and Insecurity: The Shaping of Public Perceptions of Cybercrime*, *International Review of Law, Computers and Technology*, 22 (2008), 1-2, 45-63; Alisdair A. Gillespie, *Cybercrime - Key Issues and Debates*, New York, Routledge, 2019; Marija T. Britz, *Computer Forensics and Cyber Crime: An Introduction*, London, Pearson, 2013; András Nagy, Zoltán Nagy, *Bűncselekmények számítógépes környezetben*. Budapest, Ad Librum, 2009; Zoltán Szathmáry. PhD thesis. Budapest, (PTE ÁJK), 2012; Kitti Mezei: *Current challenges of cybercrime in criminal law*. Budapest: L'Harmattan - TK JTI, 2020.

⁵ See Balázs Bartóki-Gönczi, "The relationship between search engines and freedom of expression: approach and regulatory proposals in the European Union and the United States". *Iustum Aequum Salutare*, 14 (2018), 1. 157-194; Tamás Klein: *Search Engine Providers and Internet Freedom of Expression*. In András Koltay - Levente Nyakas (eds.): *Tanulmányok a technológi- és cyberjogról einige aktuális kérdéséről*. Budapest, Média tudományi Intézet, 2018. 26-29.; and András Koltay. *The new media and the media media and the new constitutional foundations of the public sphere*. Budapest, Wolters Kluwer, 2019; Bernát Török: *Platforms and Fundamental Rights*. *Hungarian Conservative*, 1 (2021), 2. 42-47.

2. The concept of cybercrime

First, it is important to provide a brief conceptual overview of cybercrime. The term 'cyberspace' - a combination of the words cybernetics and space - was coined by the American writer William Gibson in his 1982 novel *Burning Chrome*, later made famous by his novel *Neuromancer*. Gibson used cyberspace to describe the global computer network that connects people, computers and information sources. The resulting Anglo-Saxon term cybercrime was coined after the word cyberspace. The term cybercrime is now widely used, especially in international literature, but also, for example, in the Convention on Cybercrime.⁶ However, it is important to note that any generally accepted and uniform legal definition of cybercrime has not been elaborated yet.

In the international academic scholarship, several authors, such as Jonathan Clough,⁷ Peter Grabosky⁸ and Susan W. Brenner,⁹ consider cybercrime as a generic term, with two main subcategories: one is constituted by a group of offences that can be committed exclusively with information systems (e.g. computers, their networks or other information and communication technologies - ICT). Typically, the object of such an offence is the information system. These include the so-called cyber-dependent crime (e.g. use of computer viruses, hacking, etc.). The second, broader category covers traditional crimes committed using information systems (e.g. fraud, extortion, child pornography, copyright infringement, harassment, etc.). These are similar to the case of cyber-enabled crime, where the information system serves as the instrument used to commit the crime.¹⁰

It can be concluded that the motives and aims of crimes committed in the IT environment are generally no different from those of crimes committed in real space because they can be committed for profit or damage, to obtain data or secrets, or even for sexual motives. However, political gain through manipulation and the dissemination of disinformation may also involve motives that are not relevant to traditional crimes. This is perhaps one of the reasons why these aspects have received less attention so far, even though they are becoming increasingly important for criminals in the virtual world.

The regulation of cybercrime offences in the area of criminal law should be essentially a national competence, as criminal law falls typically, but not exclusively, within the domestic jurisdiction of a state, and national criminal laws are therefore supposed to be best placed to deal with the issue. Thus, in the present study, we examine specific cases of cybercrime concerning domestic criminal law provisions. However, it is important to note that domestic law is significantly influenced in this area by the aforementioned Budapest Convention and its Additional Protocol,¹¹ as well as by relevant EU legislation (e.g. the Directive 2013/40/EU on attacks against information systems).

⁶ *Convention on Cybercrime of the Council of Europe*, signed in Budapest on 23 November 2001 and promulgated in Hungary by Act LXXIX of 2004.

⁷ Jonathan Clough, *Principles of cybercrime*, Cambridge University Press, 2015. 10-11.

⁸ Peter Grabosky, *Cybercrime*, London, Oxford University Press, 2016. 8-9.

⁹ Susan W. Brenner, *Cybercrime - Criminal Threats From Cyberspace*, Santa Barbara, CA, Praeger, 2010. 39-47.

¹⁰ Clough *ibid* (p.5) 10-11. The term *cyber-related crime* is used by the US Department of Justice when the computer is the instrument of traditional crime. At the same time, the *Budapest Convention* also refers to offences committed by using a computer. See U.S. Department of Justice, *The National Information Infrastructure Protection Act of 1996, Legislative Analysis*, 1996; Council of Europe, Explanatory Report to the Convention on Cybercrime, *European Treaty Series* - No. 185, 2001, Article 79.

¹¹ In 2003, the Budapest Convention was supplemented by a Protocol on criminalising racist and xenophobic acts committed through computer systems. In September 2017, the Council of Europe decided to draw up a second Additional Protocol to the Convention, which would include provisions for a more effective and simpler mutual legal assistance system. This system would allow for direct cooperation with service providers established in another State Party to the Convention, and searches could be carried out across borders. The

protocol will include strong safeguards and data protection requirements. The advantage of such an agreement is that it could become applicable worldwide. See Jennifer Daskal - Debrae Kennedy-Mayo: Budapest Convention: What is it and How is it Being Updated? Cross-Border Data Forum, 2 July 2020.

3. Some cases of cybercrime on online platforms

Various forms of businesses are increasingly dependent on online platforms, while individuals tend also using them even more frequently, and the coronavirus epidemic has also increased considerably the online presence of each person. Users are often unaware of the dangers of being online and the consequences of sharing information. As a result, sensitive data may be increasingly easy for unauthorised persons to obtain (e.g. through malware, phishing, and other methods developed for this purpose). Social media platforms offer an easy way for hackers to reach or map their targets. The rapidly evolving number of social media users has made online platforms attractive to cybercriminals, which means that this latter group has become a significant source of malware¹² infections, for example, affecting both individuals and businesses. The problem is growing, with these platforms being particularly well suited to malware distribution, as they tend to display more images, videos, advertisements and plugins. In addition, the nature of interaction through social networks facilitates the rapid and seamless spread of infection – an issue complicated by the tendency of social media to allow sharing user profiles across multiple platforms. One typical example of this was phishing links on Facebook Messenger, which were used to redirect victims to a page similar to YouTube. After downloading an update, users were infected with malware that was able to obtain passwords and other sensitive information.¹³

Malware is typically distributed through social media posts or messages, such as clicking on infected ads, content shared by friends (e.g. funny videos, pictures and news), and installed plugins and applications (e.g. games and tests). Often malware is sent as a message or redirected to a website. Still, drive-by downloads have also appeared, which are particularly dangerous because malware downloads itself, exploiting system vulnerabilities by embedding itself in the website or application. So-called bots are often found on social media sites, which can be used to generate automated messages and distribute malware. For example, a chatbot in Facebook Messenger can be used to send messages that may amount automatically to a criminal offence.¹⁴ In addition to malware, it is common for personal or business accounts to be hacked (so-called hacking or unauthorised access), which can lead to the user taking control of the account and gaining access to all the data associated with the account (such as credit card and other personal data). They often target accounts that are "verified". Setting up two-step authentication for user accounts is particularly important to avoid such attacks. These cases are all related to offences against the information system: the perpetrators commit a breach of the information system or data (Section 423 of the Hungarian Criminal Code), if they access the user account without authorisation by circumventing the technical measures, or carry out unauthorised data manipulation with the malicious program, or fraud is committed using the information system (Section 375 of the Hungarian Criminal Code), if, for example, transactions are carried out with the unauthorised credit card data.

One reason social media should be particularly popular with offenders is that creating a fake profile is now fairly easy. Facebook recently deleted over 5 billion fake accounts from its entire platform. Fake profiles are typically used to deceive other users, for example, by using

¹² Malware is short for malicious software, which refers to malicious programs that are commonly used to gain unauthorised access to information systems or to make changes without the user's knowledge or consent, causing damage to data. Increasingly, they are being used to gain access to confidential data that facilitates further fraud or other breaches (e.g. extortion, identity theft, etc.). For more on the different types of malware, see Kinga Sorbán, *Viruses and zombies in criminal law. Criminal substantive and procedural issues of information system and data breaches*. In Medias Res 2018/2. 376-377.

¹³ Michael McGuire: *Into the Web of Profit - Social Media Platforms and the Cybercrime Economy*, Bromium, 2019. 7-8.

¹⁴ István Ambrus: *Digitalisation and Criminal Law. Digitalisation and digitalisation*. 46.

social engineering techniques to trick them into clicking on infected links or even sharing sensitive information.

Based on the related practice of the National Authority for Data Protection and Freedom of Information, cases, where unknown persons create a fake profile on a social networking site using the user's name and photos are considered cases of suspected criminal offences. Through this fake profile, the perpetrator identifies the real acquaintances of the impersonated user and conveys messages and posts on behalf of the victim user. The aim is often to discredit the person concerned, tarnishing their reputation in the eyes of others, which can significantly damage their interests.¹⁵ Other people's personal information is also possibly used to commit crimes (e.g. using a fake profile to defraud other unsuspecting users of money or credit card details). The fake profile method often involves using other users' personal information (e.g. name and photos), which can be followed by further fraud or extortion. Examples include so-called romantic scams, where a person uses social networking sites to gain the trust of the other party with the appearance of seeking a partner and deceives or blackmails them with compromising pictures.¹⁶ In the case of fake profiles, the offence of misuse of personal data (Section 219 of the Hungarian Criminal Code) arises. In addition, data leaks pose a challenge when the mass personal data of users of social media platforms are made public, with legal consequences for data protection (see Cambridge Analytica scandal).¹⁷

Online fraud is often carried out by posting content on behalf of public figures. For example, on Twitter, a message from Elon Musk was available with the title "Dojo 4 Doge", and the scammers responded to this message and shared a link. The compelling message led to a professional website where they tried to persuade visitors to send bitcoins supposedly to Elon Musk because if they did, they would receive soon double the amount of Musk's investment. The website even said that senders above a certain amount could win the grand prize, which would be a Tesla Model S.¹⁸ In addition, social media platforms are increasingly being used to present investment and cryptocurrency scams (in the form of offers of quick returns on investment) that take advantage of the popularity of these virtual currencies and new types of investment (for example, more than 15 000 bots have been identified on Twitter in connection with cryptocurrency scams). These cases are considered traditional fraud (Article 374 of the Hungarian Criminal Code), whereby natural persons are misled by for-profit and cause damage.

Both adults and children are at risk of online sexual blackmail (sextortion), but the latter are particularly vulnerable.¹⁹ The perpetrator often gains the child's trust (for example, the adult poses as a minor and befriends the child, showing the child explicit sexual material to reduce their sexual inhibitions) and exploiting their vulnerability. This is done to gain access to sexual images or videos of the child,²⁰ followed by a phase of blackmail. The perpetrator coerces and blackmails the victim into performing sexual favours for him or sending further

¹⁵ Attila Péterfalvi - Dániel Eszteri: The criminal law protection of personal data in Hungary and the related practice of the National Authority for Data Protection and Freedom of Information. In Márta Görög - Attila Menyhárd - András Koltay (eds.). *Az Alaptörvény VI. cikkelyének érvényesülése a magyar jogrendszeren belül*. Budapest, ELTE-ÁJK, 2017. 411.

¹⁶ Réka Gyarakí: *The impact of social media on cybercrime*. Hungarian Policing 2021/2. 77.

¹⁷ Margaret Hu: *Cambridge Analytica's black box*. Big Data & Society 2020, and see Chris Hoofnagle, *Facebook in the Spotlight: Dataism vs. Personal Autonomy*, JURIST - Academic Commentary, Apr. 20, 2018, <http://jurist.org/forum/2018/04/chris-hoofnagle-facebook-dataism.php>.

¹⁸ <https://www.businessinsider.com.au/man-lost-560000-worth-of-bitcoin-elon-musk-scam-bbc-2021-3>

¹⁹ Anastasia Powell - Nicola Henry: *Sexual violence in a digital age*, London, Palgrave Macmillan, 2017. 122-124.

²⁰ For example, Snapchat is a video-sharing portal and app that is particularly popular among young people, where users can share a picture or video for a time that they set themselves, in addition to a text message.

compromising images or videos of themselves. If the request is not complied with, they threaten to share the footage already in their possession (for example, via social media), thereby gaining control over the victim.²¹

Finally, it is worth mentioning the rise of deepfake technology, which is relatively new and poses an increasingly serious challenge to society. In the case of a deepfake, an algorithm can replace the facial image in a person's video with another person's appearance, which can be deceptive to anyone. In addition to the damage caused to the individual (like revenge porn),²² deepfake can contribute to disinformation,²³ distortion of democratic decision-making and manipulation of the electoral process, eroding public trust and exacerbating societal divisions. So far, we have focused on the main characteristics and mechanisms of cyber-attacks; in the following, we will turn to the directly related aspects of freedom of expression. We will conclude with our *de lege ferenda* proposals on the principles and elements of regulation.

²¹ Europol: *Internet Organised Crime Threat Assessment (IOCTA)*. 2014. 30.

²² The Anglo-Saxon literature on this issue is dealt with in detail in Karolina Mania, *The Legal Implications and Remedies Concerning Revenge Porn and Fake Porn: A Common Law Perspective*, Sexuality & Culture 2020 and Catherine D. Marcum - George E. Higgins - Tsung Martin Tsai - Jeffrey Sedlacek, *Exploration of Prosecutor Experiences with Non-consensual Pornography*. Deviant Behavior, 42:5 2021. 646-658.

²³ Christopher Whyte: *Deepfake news: AI-enabled disinformation as a multi-level public policy challenge*. Journal of Cyber Policy 2020.

4. Freedom of expression on online platforms

4.1. Online platforms as spaces for freedom of expression

Over the past decades, and especially in recent years, platforms in virtual space have become the primary arena for the clash of political opinions. As more and more voters are using online platforms intensively, and as the content, they wish to disseminate reaches their target audiences more quickly and effectively through these channels, they have become increasingly popular for political communication. Political parties and candidates approach their voters primarily in virtual space, and citizens reflect on the views shared with them via virtual channels of communication.²⁴ Private opinions are also confronted mainly on the Internet, where they can be disseminated with unprecedented speed and efficiency and often anonymously commented on.²⁵

These processes were particularly accelerated during the pandemic when curfews and contact restrictions forced virtually everyone to increase their online presence.²⁶ In addition, some of the measures taken in the context of the epidemic explicitly affected certain forms of expression; for example, assemblies were banned in many places or allowed only within strict limits and with a restricted number of participants. Under these circumstances, it was necessary to organise electoral campaigns or maintain political discourse for citizens interested in participating. This was only possible, or at least predominantly likely, through online platforms, whose role was thus even more important.

The virtualisation of political communication raises several political science and sociological questions, and in this paper, we will deal with the political aspects of this problem. Only a minority of users are aware of these platforms and the risks involved in using them, mainly due to the malicious cyber activities outlined above.²⁷ Furthermore, in most cases, by making their views public, the persons expressing them are unaware they are becoming parties to a legal relationship involving at least three parties.²⁸ The platform operator provides a platform for individuals to express and share their views with other users and, in this context, to exchange views. The operator is primarily responsible for the proper operation and continued availability of the platform but, to a limited extent, must also be responsible for the content shared on the online platforms it maintains. On the other hand, individual users of the service may be primarily responsible for possible infringements through their own communications, for example, by publishing hate speech.²⁹ In addition, we must also consider the possibility of third parties becoming involved in the legal relations relating to the online exchange of opinions, such as those who engage in phishing or seek to distort democratic discourse.

²⁴ Bernát Török: *The dynamics of the protection of freedom of expression*. Hungarian Law 2017/12.

²⁵ The right to freedom of expression and the use of encryption and anonymity in digital communications - Submission to the United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression by the Association for Progressive Communication (APC).

<https://www.ohchr.org/Documents/Issues/Opinion/Communications/AssociationForProgressiveCommunication.pdf>

²⁶ Andrea Matwyshyn: *Cyber Harder*. 24 *Boston University Journal of Science and Technology Law*, 450th (2018).

²⁷ Joe Burton, *Cyber-Attacks and Freedom of Expression: Coercion, Intimidation and Virtual Occupation*, *Baltic Journal of European Studies*, 9 (2019), 3. 16-133.

²⁸ Pavlina Pavlova (2020) 'Human-rights based approach to cybersecurity: addressing the security risks of targeted groups. *Peace Human Rights Governance*, 4 (2020), 3. 391-418. DOI: 10.14658/pupj-phrg-2020-3-4.

²⁹ Bernát Török: *Media law standards for the content of hate speech*. Law Bulletin 2013/2.

4.2. Disinformation

Cyberattacks affect the development of the democratic discourse on platforms in two ways: on the one ³⁰hand, they distort the marketplace of opinions at the system level, and on the other hand, can act as a disincentive for individuals to engage in public debate.³¹ In this subsection, we first address the most important systemic risk, the creation and dissemination of disinformation.

According to the European Commission, disinformation is "information that is verifiably false or misleading, is created, published or disseminated for commercial advantage or with intent to deceive, and is likely to harm the public interest."³² In 2018, the European Commission also set up a group of experts to explore the mechanisms linked to the spread of fake news and online disinformation.³³ There can be four main strands to the disinformation phenomenon, and new technologies can contribute to them. On the one hand, by generating fictitious user profiles and articulating certain positions through them, cyber activity can bring to the fore aspects of the discourse that there is no real social need to discuss.³⁴ This is also facilitated by the fact that real persons often express their views on online platforms under pseudonyms or even anonymously, so users cannot distinguish between comments and positions representing the real person and fictitious ones.³⁵

Another alternative may be to exaggerate or even relativise the importance of the positions already present, which can influence public opinion because one increasingly draws conclusions about the current state of public opinion based on the communications one sees on online platforms.³⁶ Thus, if we perceive that most participants support a particular position, in some cases as a result of manifestations generated in whole or in part by cyber tools, we will assume that the public mood is the same. It is a sociological issue but also has constitutional implications through the influence of the electorate's will. The creation of this subjective feeling can have a considerable impact on public discourse and even on the outcome of individual elections.³⁷ The most organised forms of this manipulation of public opinion are troll farms, which are set up to influence the political process and decision-making by disseminating false information.³⁸

³⁰ Gregory T. Nojeim: Cybersecurity and Freedom on the Internet.

https://jnsplp.com/wp-content/uploads/2010/08/09_Nojeim.pdf

³¹ Noha Fathy: Freedom of expression in the digital age: enhanced or undermined? The case of Egypt. *Journal of Cyber Policy*, 3 (2018), 1. 96-115.

³² Joint Report of the European Commission and the High Representative of the Union for Foreign Affairs and Security Policy on the implementation of the Joint Action Plan against disinformation: <https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:52019JC0012&from=EN>

³³ <https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation>

³⁴ Holly A. Garnett - Toby S. James: *Cyber Elections in the Digital Age: Threats and Opportunities of Technology for Electoral Integrity*. <https://doi.org/10.1089/elj.2020.0633>

³⁵ The fight against disinformation and the right to freedom of expression Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, PE 695.445 - July 2021.

[https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU\(2021\)695445_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf)

³⁶ Balázs Bartóki-Gönczy: The relationship between search engines and freedom of expression. Approach and regulatory proposals in the European Union and the United States. *Iustum Aequum Salutare* XIV (2018), 1. 157-194.

³⁷ Kevin M. Caramancion: An Exploration of Disinformation as a Cybersecurity Threat. 2020 3rd International Conference on Information and Computer Technologies (ICICT). (2020), 440-444. and László Kovács - Csaba Krasznay. Nation and Security: Security Policy Review 2017/3 3-15.

³⁸ For more on troll farms, see Jamieson Kathleen Hall, *Cyberwar: How Russian Hackers and Trolls Helped Elect a President: What we Don't, Can't, and Do Know*. Oxford, Oxford University Press, 2018.

The third related trend is using fictitious profiles on online platforms to present facts that have no basis in reality. The extremely high risk of such disclosures is particularly striking, given the structure of modern public opinion.³⁹ Even the most astonishing fake news reaches a wide range of users very quickly and spreads the fastest, while the content that exposes its falsity is of much less interest. A big fake news story can therefore have a significant negative impact on the perception of specific public figures or even on assessing certain issues on the political agenda. The creation of fictitious profiles is subject to criminal prosecution as a misuse of personal data, as well as establishes civil liability as a violation of right to image. Moreover, the managing structure of a particular website, as well as the funder of certain advertisements remain hidden for the public which enhance the probability of spreading disinformation through these manners. The digital services act addresses these issues with the duty on platform providers to remove any illegal content from their spaces; the annual monitoring by the member states could enforce the proper execution of this mandate. Nevertheless, one should expect, that the issue of fictitious profiles will not disappear from the agenda for a longer period.

Finally, we must also take into account the form of disinformation, when factually true statements appear out of context, distorted in such a way that it is suitable to mislead consumers of the content, to create an impression that is not actually true, without stating factually false facts.⁴⁰

In addition to the systemic negative consequences, we will now look at the threats at the individual level that may deter individuals from engaging in online public debate or affect unsuspecting individuals who express their views in virtual spaces.

³⁹ Allison O. Larsen: Constitutional Law in an Age of Alternative Facts. 93 *New York University Law Review*, (2018), 2. 175; 178.

⁴⁰ Ari E. Waldman, The Marketplace of Fake News. 20 *Journal of Constitutional Law*, 4 (2018), 101-105;

4.3. Cyber threats to freedom of expression through virtual platforms

4.3.1. Data protection concerns

At the individual level, one main motivation for platform cyber activity is illegal data acquiring.⁴¹ The purpose of this can be twofold: on the one hand, to profile the individuals concerned, even based on their views, and on the other hand, to misuse the personal data acquired (for example, for financial benefit) without any political motive. The data protection challenges related to the fate of the personal data of platform commentators can be grouped into four main categories.

On the one hand, the identity of the natural persons behind cyber-attacks is often untraceable or very difficult to trace. So it is not transparent who is getting hold of our personal data that is not sufficiently protected.⁴² The result of this lack of transparency is that we can completely lose control over the fate of our personal data, and often, enough information to build a personality profile can end up in hands we don't even know.

On the other hand, the fact that people with opaque backgrounds can gain access to users' personal data is not a major issue in itself, but because it is unpredictable what the phishers intend to do with the personal data they have unlawfully processed.⁴³ This is particularly important given that when we express our opinions on online platforms, we often take positions on very sensitive issues where anonymity. At the same time, a risk, can also be guaranteed.⁴⁴ Thirdly, cyber tools can also be used to recognize anonymous commentators, identify stakeholders who have given their name, and reveal personal data they did not wish to share. However, in many cases, we are not simply talking about personal data but about particularly sensitive personal data, which is why many people prefer to stay away or refrain from online discourse at a time when more and more of the dialogue on issues of public interest is being shifted to these platforms.

The fourth cyber threat to the integrity of the platforms is the organisation of the acquired data of the users of these platforms into databases, which can provide cybercriminals with a picture not only of the individual but also of their place in the wider social environment.

⁴¹ Elizabeth F Judge - Michael Pal: Voter Privacy in the Age of Big-Data Elections. 58 *Osgoode Hall Law Journal*, (2021), 1. 2.

⁴² Lyria B. Moses, Recurring Dilemmas: The Law's Race to Keep Up with Technological Change, *Journal of Law Technology & Policy* (2007), 239; 274-275.

⁴³ Ira S. Rubinstein: Voter Privacy in the Age of Big Data. 5 *Wisconsin Law Review* (2014), 861.

⁴⁴ András Koltay - Levente Nyakas: *Studies on some current issues of technology and cyber law*. Edited by Tamás Klein. Institute of Media Studies, 2018. 18-19.

4.3.2. Obstructing democratic discourse

As a main weightful actor rather than data protection issues, the individual's situation in an increasingly virtual democratic space is becoming more and more difficult for individual citizens to navigate in an increasingly complex and opaque marketplace of opinions.⁴⁵ This is something that a significant proportion of platform users are aware of, with the decreasing reliability of information sources and the increasing manipulation of political communication.⁴⁶ This reinforces apolitical tendencies in society, further hindering the development of an inclusive democratic discourse.

A further severe difficulty around platforms is that their operators often moderate the content of public discourse by removing undesirable content, at most according to their internal rules.⁴⁷ Such interference is usually always aimed at a consensual goal, such as curbing hate speech, and protecting human dignity or the dignity of individuals or well-defined social groups. However, there are already significant differences in interpreting these concepts, and many people feel that their expressions have been unfairly removed from the opinion market by platform operators. Moreover, the operation of platforms is poorly regulated by law, and the background of their operators is often not very transparent, so the framework of political discourse and often the limits of the individual's freedom of communication are decided by the platform operators, i.e. private actors with no public authority,⁴⁸ and often without the moderators themselves or the interest groups behind them being identifiable.

⁴⁵ Rebecca Green: Counterfeit Campaign Speech. 70 *Hastings Law Journal*, (2019). 1445.

⁴⁶ Elizabeth F Judge - Amir M Korhani: A Moderate Proposal for a Digital Right of Reply for Election-Related Digital Replicas: Deepfakes, Disinformation, and Elections. *SSRN*, id: 3827249

⁴⁷ Katalin Parti: Fighting illegal content online.

http://www.okri.hu/images/stories/KT/KT_49_2012/004_parti.pdf

⁴⁸ Elizabeth F Judge - Amir M Korhani: Digital Information Equality, Disinformation, and Elections. 19 *Election Law Journal*. (2020), 240.

5. The case law of the European Court of Human rights

The ECtHR is well aware of the importance of the internet in relation to freedom of expression. It declared: “The Court notes at the outset that user-generated expressive activity on the Internet provides an unprecedented platform for the exercise of freedom of expression.”⁴⁹ Another recurring stance laid out in most of the relevant case law concerns the increased general access to news, and the platform as a source of dissemination. The ECtHR held that “the Internet plays an important role in enhancing the public’s access to news and facilitating the dissemination of information in general.”⁵⁰ Several valid reasons were provided why the online marketplace of ideas ought to be examined and regulated differently than other tools of communication. Considering the possible effects, the ECtHR found that audiovisual media have a more immediate and powerful effect than print media.⁵¹ The Jersild judgement explains that “The audiovisual media have means of conveying through images meanings which the print media are not able to impart.”⁵²

By itself, this is definitely not a negative phenomenon. It simply means that we can send and receive information quicker than ever before. On the other hand, the ECtHR is also right that there is immense risk in all of this, because even if a post is factually incorrect, or intentionally or inadvertently misleading, it will remain online long enough to be seen by a wide audience. Hence, “The risk of harm posed by contents and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press.”⁵³ It can be argued that from the standpoint of a state, this principle can especially serve as a legitimate aim to intervene in people’s freedom of expression by passing stricter regulations during a public health emergency or during the period of a war.

There is another matter we must touch upon, which is the duty and the responsibilities of media portals. Portals which provide a forum for the exercise of the right of expression, enabling the public to impart information and ideas, “must be assessed in the light of the principles applicable to the press.”⁵⁴ The question is whether such portals can be classified as the publishers of thirdparty content. The ECtHR established in a notable case that these sites are “not publishers of the comments in the traditional sense”. But this does not mean that they have no responsibility at all: “Internet news portals must, in principle, assume duties and responsibilities.”⁵⁵ Therefore, internet platform providers’ duties differ greatly from those of traditional publishers, and include “(a) large news portal’s obligation to take effective measures to limit the dissemination of hate speech and speech inciting violence”. However, there is a limitation: this “can by no means be equated to private censorship”.⁵⁶ In summary: the platform holder bears responsibility not because it is the publisher, but upon consideration of three conditions: if (a) the publication of the comments is in its financial interest, and (b) it increases the page’s popularity, and (c) no “notice and take down” system or anything similar is in place that would result in the immediate removal of the offensive comments.⁵⁷ It is therefore in the interest of these portals to create a moderated online environment, if they wish to avoid being punished for their unwillingness to take down harmful comments. The platforms

⁴⁹ Delfi AS v. Estonia (ECtHR, June 16, 2015, no. 64569/09). § 110.

⁵⁰ Delfi AS v. Estonia (ECtHR, June 16, 2015, no. 64569/09). § 133.

⁵¹ Monnat v. Switzerland (ECtHR, December 6, 2006, no. 73604/01). § 68.

⁵² Jersild v. Denmark (ECtHR, September 23, 1994, no. 15890/89). § 31.

⁵³ Editorial Board of Pravoye Delo and Shtetel v. Ukraine (ECtHR, May 5, 2011, no. 33014/05). § 63.

⁵⁴ Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary (ECtHR, February 2, 2016, no. 22947/13). § 61.

⁵⁵ Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary (ECtHR, February 2, 2016, no. 22947/13). § 61.

⁵⁶ Delfi AS v. Estonia (ECtHR, June 16, 2015, no. 64569/09). 57.

⁵⁷ Delfi AS v. Estonia (ECtHR, June 16, 2015, no. 64569/09). § 162.

provided to user-generated content also have an important role because, as the ECtHR determined, they foster the emergence of citizen journalism, and therefore content ignored by traditional media can prevail.⁵⁸

⁵⁸ Cengiz and Others v. Turkey (ECtHR, December 1, 2015, no. 48226/10 and 14027/11). § 52.

6. Recommended regulatory solutions

Cyber-attacks may occur via online platforms because these platforms are extremely under-regulated, with no legal codes cover the related liability.⁵⁹ It is not clear what the legal obligations of the platform operator or user are nor what behaviour they are obliged to adopt to prevent cyber-attacks. An excellent example to demonstrate this global uncertainty may be the decision of the French Constitutional Council in June 2020, which annulled a law under which a platform operator could be fined a substantial amount of money if it did not remove the infringing content from its platform within 36 hours after it was published.⁶⁰ This decision also showed that no clear legal requirements have been set from the various participants in the exchange of ideas on online platforms.⁶¹

There is a broad consensus that legislation needs to take action in this area, but it is questionable what direction it should take to improve the situation. At the European level, a draft specific regulation on digital services is currently being discussed as part of the legislative package on digital services. This proposal would bring several innovations for platform operators. In essence, it would not generally sanction platform providers for failing to remove illegal content shared on their platforms. Still, it would oblige these stakeholders to remove the communication in question immediately if they become aware of the illegality. While this is not yet fundamentally revolutionary, as it is all in line with existing legal requirements and practice, it does impose new obligations on the most prominent platform providers: they must disclose the principles of their artificial intelligence-based algorithms that analyse communications and how they decide to remove certain content from their platforms. This will therefore provide some transparency to users on how their comments are judged and may reduce one of the factors that can deter many from commenting on online platforms. However, neither the Digital Services Act nor any other currently known draft legislation addresses how to alleviate the pressure on platforms to promote the safety of those who comment on them.

In our view, the starting point in addressing the challenges to freedom of expression is that only an integrated approach and a combination of instruments can achieve meaningful results in this area.⁶² Several technological, economic, personal and legal conditions would be necessary for cleaning the discourse on platforms from manipulated content and cyber-attacks. One need to talk about technological requirements because One need to constantly improve the IT solutions that can prevent malicious interference from cyberspace in the operation of platforms. It is also necessary to continue this reflection from an economic point of view because cybercrime affecting platforms is primarily motivated by such factors, which we must identify and counteract. On a personal level, an essential prerequisite for platform protection is to have professionals who can both identify the main challenges and work out the best ways to tackle them and who can also be involved in the day-to-day defence with the technologies at their disposal.

The legal regulation should reflect the pressure on the platforms, taking into account the above aspects. In our view, one way of doing so could be to share the responsibility for

⁵⁹ Philip N. Howard: *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. New York: Oxford University Press, 2018.

⁶⁰ Décision n° 2020-801 dc du 18 juin 2020.

<https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm>

⁶¹ Boldizsár Szentgáli-Tóth: Keep your eyes on Paris - or against hate speech with or without fines? *Legal Forum Press and Media Blog*, (June 2021) <https://www.jogiforum.hu/blog-sajto-es-mediajog-blog-13/2021/06/30/vigyazo-szemetek-parizsra-vessetek-avagy-penzbirsaggal-vagy-anelkul-a-gyuloletbeszed-ellen/>

⁶² Jamie Lund: Correcting Digital Speech. 19 *UCLA Entertainment Law Review*, (2012), 170; 186.

protection between the platform operator and the user. For operators, there should be requirements on what security measures they must take and what technological solutions they should use to prevent cyber attacks. Suppose platform providers fail to comply with this obligation within a timeframe that gives them sufficient time to prepare. In that case, they could be fined and, ultimately, forced to cease operating the platform. At the same time, the responsibilities of users should be clarified: in which cases are users of platforms expected to act prudently, and what is the unprudent behaviour in virtual space that should be carried out without the user having to bear the consequences of their carelessness? This could be the case, for example, where a user makes personal data voluntarily available to an unreasonably large number of people, which are in no way related to the use of the platform or the opinions shared with others through the platform.

Just as the legal status of the platform operator and its users in the context of online commenting is generally not elaborated, the same is not true for protecting against cyberattacks. A more informed and effective response to this difficult-to-identify threat can only be expected if the respective actors may foresee clearly their legal responsibilities in this area. Sanctions should be applied gradually and only as a last resort if no other means of enforcing risk mitigation can be found.

In the European space, there may also be other legal instruments to strengthen cooperation between the Member States to curb cross-border cybercrime. As the fundamental characteristic of these offences is their cross-border nature, cooperation at the European level is of fundamental interest for crime prevention. On the one hand, there is a need for broader cooperation than is currently the case to obtain electronic evidence. On the other hand, data related to this type of crime should be made available to all interested authorities and researchers working on related issues.

Concluding remarks

Current trends in cybercrime have been addressed by several authors, and changes in freedom of expression are often analysed in the literature. Yet few attempts have been made to date to outline the legal implications of the challenges of freedom of expression in cyberspace by focusing on the intersection of these two seemingly distant disciplines. We consider this a serious shortcoming, especially in the light of the fact, that an increasing part of the public discourse on public affairs is being shifted to online platforms, owing to the physical isolation resulting from the epidemic.

Another meaningful issue derives from the fact, that the existing and still insufficient legal framework including the ECtHR case law concentrates only on platform providers and users. At the same time, there are few studies and practical experience that focus on the possible legal means of collective defence against external actors. We have proposed some basic principles and institutions for this legal concept, which needs to be developed, emphasising the sharing of responsibility between platform operators and users. We believe that recognising the true significance of cyber-attacks impact on democratic discourse, and developing the legal environment accordingly, could help to reduce the manipulative nature of political communication on platforms, thereby affecting all forms of democratic participation and the daily lives of all, or at least many, citizens.

In this paper, we have not sought for providing definitive answers to these prospective objectives but to draw attention to some new aspects of the dilemmas already discussed in the literature. However, further extensive professional discussion will be needed to develop a long-term approach to the challenges of freedom of expression, and we hope that our suggestions have contributed to set the direction for this discourse.