TIINA PAJUSTE (ED.)

# Specific Threats to Human Rights Protection from the Digital Reality

## International Responses and Recommendations to Core Threats from the Digitalised World

EU COST Action – CA19143 – Global Digital Human Rights Network 2022

„All human beings are born free and equal in dignity and rights.“

Art. 1, sentence 1, Universal Declaration of Human Rights (1948),

EU COST Action – CA19143 – Global Digital Human Rights Network

# Specific Threats to Human Rights Protection from the Digital Reality

## International Responses and Recommendations to Core Threats from the Digitalised World

**edited by Tiina Pajuste**
TALLINN UNIVERSITY

**Cite as:** Tiina Pajuste (ed.), Specific Threats to Human Rights Protection from the Digital Reality (Tallinn: Tallinn University, 2022)



**Publisher:** Tallinn University
Narva mnt 25, 10120 Tallinn, Estonia
www.tlu.ee

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

TALLINN UNIVERSITY

# Introduction

Digital technologies are bringing widespread benefits to people and communities across the globe. The UN has acknowledged that digital technologies can make the world fairer, more peaceful, and more just. Reaching the 17 Sustainable Development Goals can all be aided by digital means and advances.[1] Digital technologies can have a very positive effect on many human rights. E.g., AI-aided technologies have helped diagnose diseases and save lives (helping ensure the right to health) and virtual learning has provided access to education to a multitude of students who would otherwise be excluded. However, digital technologies can also bring with them a multitude of problems and potential threats to human rights protection, especially if they are misused or the necessary protective actions are not taken. There is a pressing need for adequate protection by governments (vertical approach) and online stakeholders (horizontal approach). As technological developments take place faster than regulation than can catch up, the guidelines and rules that have been adopted have not provided solutions to many of the issues, including the human rights' implications of digital activities.

The most recent UN Human Rights Council report on the right to privacy in the digital age engages with numerous threats to privacy and highlights how digital tools can expose people to new forms of monitoring, profiling and control. It focuses on three trends in state actions in relation to the right to privacy: '(a) the widespread abuse of intrusive hacking tools; (b) the key role of robust encryption in ensuring the enjoyment of the right to privacy and other rights; and (c) the widespread monitoring of public spaces. The report highlights the very real and encroaching risk of creating systems of pervasive surveillance and control that may eventually choke the development of vibrant, prosperous and rights-respecting societies'.[2]

There are also threats that can lead to increased discrimination. For example, AI and algorithms can replicate or even amplify human and systemic bias, if they are based on data that is lacking diversity. This may be amplified by a general lack of diversity in the technology sector. Additionally, not everyone has the same opportunity to benefit from digital technology. Often it is women, the elderly or other vulnerable groups that remain disconnected, or, if they have access to internet, it may be less meaningful due to lack of digital skills.

It is important to be aware of the potential threats in order to fully harness the positive potential of digital technologies. This report hopes to contribute to this goal. It does not claim to represent a full list of relevant threats and challenges, but instead aims to highlight some core problems that can arise in the digital realm. The report is divided into five parts: (a) overarching issues that impact human rights in general, (b) threats that are connected to technology/technological advancements, (c) threats connected to disinformation and manipulation online, (d) threats connected to privacy and freedom of expression, and (e) threats in relation to vulnerable groups.

[1] UN, 'The Impact of Digital Technologies', available online at: https://www.un.org/en/un75/impact-digital-technologies .

[2] UN Human Rights Council, 'The right to privacy in the digital age: Report of the Office of the United Nations High Commissioner for Human Rights', UN Doc. A/HRC/51/17, 4 August 2022, para 3.

The amplitude of different threats makes it essential that these technologies are reined in by effective regulation based on international human rights law and standards. The initiatives taken on both regional and global levels are noted in this report in relation to each threat that is addressed.

Each report section concludes with recommendations to different stakeholders for policies and activities that are aimed at minimising threats to human rights protection from the digital environment. The report is a compilation of contributions from twenty-two researchers from the Global Digital Human Rights Network.

**Tiina Pajuste**

*Professor of International Law and Security Studies at Tallinn University*

*Leader of Working Group 1 of Global Digital Human Rights Network*

# List of authors

**Artūrs Kučs**, University of Latvia

**Barbora Baďurová**, Matej Bel University

**Birgit Schippers**, University of Strathclyde

**C H Powell**, University of Cape Town

**Cristina Elena Popa Tache**, Institute of Legal Research of the Romanian Academy

**Eva Lievens**, Ghent University

**Igor Serotila**, American University of Moldova

**Irena Barkane**, University of Latvia

**Jukka Viljanen**, Tampere University

**Konstantinos Kouroupis**, Frederick University

**Maria Biliri**, University of Athens

**Marijana Mladenov**, University Business Academy in Novi Sad

**Nikolas Thomopoulos**, University of Surrey

**Oscar Puccinelli**, Rosario National University

**Pinelopi Troullinou**, Trilateral Research

**Saulius Stonkus**, Vilnius University

**Šejla Maslo Čerkić**, OSCE Mission to Bosnia and Herzegovina

**Skirgailė Žalimienė**, Vilnius University

**Tiina Pajuste**, Tallinn University

**Vesna Crnić-Grotić**, University of Rijeka

**Violeta Beširević**, Union University (Belgrade)

**Vygantė Milašiūtė**, Vilnius University

# Table of contents

# Overarching issues in the digital realm

**Vygantė Milašiūtė**
VILNIUS UNIVERSITY

**Violeta Beširević**
UNION UNIVERSITY, BELGRADE

**Oscar R. Puccinelli**
ROSARIO NATIONAL UNIVERSITY

**Cristina Elena Popa Tache**
INSTITUTE OF LEGAL RESEARCH OF THE ROMANIAN ACADEMY

**Barbora Baďurová**
MATEJ BEL UNIVERSITY

## Introduction[3]

While the Internet offers numerous possibilities for exercising human rights, it also has certain features which make it a source of risk and threats for human rights. Some of these threats manifest themselves only in very specific contexts, whereas others are of a general and overarching nature.

Some of those general threats are of a legal nature as they stem from uncertainty, limited scope, or divergent interpretations of law. Such is a threat of social platforms evading responsibility for human rights violations as under constitutional law of many jurisdictions they are not considered to be human rights bearers. Another threat of a similar nature is the uncertainty of whether another category of private actors, namely, investors, are bound by human rights standards when they are involved in activities in the digital realm.

Some other threats are legal in a sense that they could be addressed by legal means, but legal standards are not yet developed because the phenomenon causing a human rights problem is not yet understood to a sufficient extent. Such is a threat of cancel culture and social engineering in the context of elections where the harm of manipulated voting is obvious but how exactly the prohibited acts should be defined or how they could be prevented is an open question. Similarly, a question to what extent the risk of addiction to the Internet could provide entitlements to health care services or lead to limitations of online time is linked to a question whether mental health research provides sufficient basis for any such conclusion.

Ultimately, there is also a problem of the insufficiency of the legal approach to tackling digital human rights issues. Not only is there a need for concrete evidence to shape a legal response to Internet-

[3] Introduction written by Vygantė Milašiūtė.

related human rights threats, but also there is the inherently multidimensional nature of compliance with human rights in the digital realm, which requires a multidimensional response, including in the context of awareness raising and training. Law should be complemented by ethics and the training on cyber ethics should be provided along with training on digital human rights law.

## Eroding constitutional protection online[4]

### Description of the threat

It is a common understanding that social platforms affect internet users' human rights. Online intermediaries routinely interfere with users' right to privacy, freedom of speech, the right not be discriminated and intellectual property rights, to name just the most vulnerable. They enjoy practically unchecked power to collect private data, block, filter, censor, manipulate, or expose vulnerable groups, particularly children, to contents that can have a dangerous impact on their mental or even physical health and development.[5]

The current tendency to protect human rights in the digital sphere is reflected in states' efforts to control social platforms through administrative and legislative regulations, international soft-law instruments applicable to transnational corporations, such as UN Guiding Principles on Business and Human Rights, and social platforms' self-regulations. For example, Facebook defines binding policies, norms, and standards of behavior that users must follow when participating in the Facebook community. It also enforces these policies, meaning that Facebook regulates and adjudicates conflicts between itself and its users or between users themselves.[6] In Busch's view, Facebook amounted to an almost state-like institution in itself, with many of the major characteristics of a developed political system.[7] Consequently, social platforms exercise a power akin to state power which represents the specific challenge for constitutional democracies.

However, while a state faces constitutional responsibility for human rights violations, social platforms in most jurisdictions still evade this responsibility. This is because the law generally regards the state as the primary, if not sole, bearer of duties correlative with rights.[8] Rights do not regulate the relations between private parties whose autonomy should stay free from the compulsory regime created by the constitutions. In other words, constitutional rights apply only in relations between the government and an individual (vertically) but not between private parties (horizontally). Although the constitution generally does not sanction private deprivations of constitutional rights in most jurisdictions, notable exceptions do exist. The horizontal effect of human rights, either direct or indirect, has been recognized in several jurisdictions, most notably Ireland, South Africa, Germany, Canada, the European

[4] Report section written by Violeta Beširević, Professor of Law, Union University Law School Belgrade.

[5] For more see Nicola Lucchi, "Internet Content Governance and Human Rights", 16 *Vanderbilt Journal of Entertainment and Technology Law* 809, 2014.

[6] For more see Moran Yemini, "Missing in 'State Action': Toward A Pluralist Conception of The First Amendment", 23 *Lewis & Clark Law Review* 1149, 2020.

[7] Thorsten Busch, Fair Information Technologies: The Corporate Responsibility of Online Social Networks as Public Regulators 71 (2013) (unpublished dissertation) (on file with the University of St. Gallen), https://www.alexandria.unisg.ch/228863/.

[8] See Violeta Beširević, "Uhvati me ako možeš": o (ne)odgovornosti transnacionalnih korporacija zbog kršenja ljudskih prava [ "Catch Me If You Can": Reflections on Legal (Un)Accountability of Transnational Corporations for Human Rights Violations], *Pravni zapisi*, (2018), no.1, pp.21-42.

Union, and several Latin American states. The emerging trend, although still sporadic, is its application in digital sphere.

## Developments in relation to the threat

The European Union, a transnational political entity of a constitutional nature, is at the forefront of addressing the responsibility of social platforms for human rights violations in the digital surrounding. First, the EU will soon adopt the Digital Services Package, composed of the Digital Market Act and the Digital Service Act, which should be operable across the EU from 1 January 2024 the lattes. The Digital Service Act is essential as it provides measures for securing users' human rights online, including enhanced supervision and enforcement by the European Commission when it comes to the responsibility of very large online platforms.[9] Second, in recent years the Court of Justice of the European Union (CJEU) has taken an active approach to the human rights challenges arising from digital technologies. Several cases, including *Google Spain*[10], *Scarlet*[11], *Netlog*[12], *Digital Rights Ireland*[13], and S*chrems I*[14], testify that the CJEU is not ready to tolerate the rise of social platforms' power at the expense of users' human rights.[15] A decisive step for the CJEU's reactive approach has been the legally binding nature of the EU Charter of Fundamental Rights, confirmed in the Lisbon Treaty, and its inclusion in the EU constitutional framework.[16]

Regarding national jurisdictions, Germany took an especially proactive approach, partially thanks to the judge-made *Drittwirkung* doctrine, established in 1958. Broadly speaking, the German Federal Constitutional Court (FCC) ruled that the constitutional values embedded, for example, in human rights, do not only oblige the state but private parties as well, meaning that fundamental rights have an indirect horizontal effect on the relations between private parties. In 2019, the FCC, although not in a firm voice, extended the application of the *Drittwirkung* doctrine to the digital realm. Referring to social networks and assessing their rising powers, in *obiter dictum*, the FCC established that "the biding effects of fundamental right on private actors can ultimately be close, or even equal to its binding effect on the state."[17] However, starting from the *Drittwirkung* doctrine, the Federal Court of Justice (FCJ) recently concluded that Facebook is not big enough to be in the state's position and that, consequently, constitutional rights do not limit its activities.[18] Yet, important footnote should be added here. Although the Court found that Facebook enjoys certain rights in the digital sphere, including freedom of expression and the right to conduct business, the FCJ held that it could not regulate users'

[9] See at https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2545

[10] C-131/12, Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, [2014] ECLI:EU:C:2014:317.

[11] C-70/10, Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM), [2011] ECLI:EU:C:2011:771.

[12] C-360/10, Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV,[2012] ECLI:EU:C:2012:85.

[13] Joined Cases C-293/12 and C-594/12, Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others,[2014] ECLI:EU:C:2014:238.

[14] C-362/14, Maximillian Schrems v. Data Protection Commissioner, [2015] ECLI:EU:C:2015:650.

[15] See in Giovanni De Gregorio, "Digital Constitutionalism across the Atlantic", (2022), *Global Constitutionalism,* Vol.11, No. 2., pp. 297–324.

[16] *Ibid*., p. 305.

[17] See FCC Order, 6 November 2019, -No.1 BvR 16/13 (Right to be forgotten I), para.88.

[18]See FCJ, III ZR 179/20 and III ZR 192/20 from July 29, 20221.

actions without considering their rights. Put differently, Facebook can develop its own rules and enforce them, but not without respecting procedural requirements (similar to a state), including informing users about its actions before and after their enforcement and respecting their right to redress.[19]

In sharp contrast to this view stands a firm understanding of the US Supreme Court that constitutional rights are shields only against the state. Under its state action doctrine, the US Constitution applies only to governmental conduct and does not extend to the behavior of private parties. Those who benefit the most from this approach are online platforms. Based on the state action doctrine, the US courts have repeatedly rejected free-speech-related claims against social platforms.[20]

### Minimization of the threat – recommendations

Individual rights and state duties are not exclusively tied either normatively or historically. John Locke's rights theory, which inspired the American Declaration of Independence, never saw rights as creating duties only for government, particularly concerning the negative duty not to harm.[21] The fact that social platforms are corporate entities and are in a position to cause enormous harm to human rights supports the need to recognize them as human rights duty-bearers. Since social platforms enjoy great power and no constitutional responsibility, it is high time to challenge the standard public/private division that still dominates constitutional law across the globe.

## Internet addiction (problematic usage of the Internet)

### Description of the threat

The umbrella term problematic usage of the Internet (PUI) used by mental health researchers encompasses all potentially problematic Internet related behaviours, including those relating to gaming, gambling, buying, pornography viewing, social networking, "cyber-bullying," "cyberchondria" among others. PUI may have mental and physical health consequences.[22] Essentially, PUI is an Internet addiction, and it is a threat to health. In human rights terms, this threat can affect the right to health, right not to be discriminated against on the basis of Internet addiction diagnosis, and rights of specific vulnerable groups such as rights of the child. In certain contexts, this threat also affects consumer rights to safety of products and services.

[19] For more see Matthias C. Kettemann, Torben Klausa, Regulating Online Speech: Ze German Way, LAWFARE, 20 September 2021, https://www.lawfareblog.com/regulating-online-speech-ze-german-way

[20] Yemini, *supra note* 12, p.1168.

[21] Steven R. Ratner, "Corporations and Human Rights: A Theory of Legal Responsibility", 111 *Yale Law Journal* 443, 2001.

[22] NA Fineberg, Z Demetrovics, DJ Stein, K Ioannidis, MN Potenza, E Grünblatt, M Brand, J Billieux, L Carmi, DL King, JE Grant, M Yücel, B Dell'Osso, HJ Rumpf, N Hall, E Hollander, A Goudriaan, J Menchon, J Zohar, J Burkauskas, G Martinotti, M Van Ameringen, O Corazza, S Pallanti, SR Chamberlain, Manifesto for a European research network into Problematic Usage of the Internet, European Neuropsychopharmacology, Volume 28, Issue 11, 2018, pages 1232-1246, ISSN 0924-977X, https://doi.org/10.1016/j.euroneuro.2018.08.004.

## Developments in relation to the threat

This threat is to some extent addressed in the context of digitalization (digital market, digital services, digital rights), human rights, consumer rights.

### United Nations

In 2021, the Committee of the Rights of the Child adopted General Comment No. 25 (2021) on children's rights in relation to the digital environment. It generally recognised health risks related to the use of the Internet.

### G20

In 2021, G20 digital ministers adopted High Level Principles for Children Protection and Empowerment in the Digital Environment. It mentions the need to provide access to and make the public aware of legal, psychosocial, or therapeutic services available to children requiring assistance as a result of activities or action in the digital environment. It also prescribes raising awareness of online commercial practices that may cause children harm.

### Council of Europe

The Council of Europe adopted Guidelines to respect, protect and fulfil the rights of the child in the digital environment[23] in 2018 and developed a Handbook for policy makers on the rights of the child in the digital environment[24] in 2020. Some health risks related to the use of the Internet and the problem of premature exposure to the Internet are recognised.

### European Union

In 2022, the Commission proposed a Declaration on Digital Human Rights and Principles. A number of provisions may be relevant for preventing Internet addiction, but such a threat is not specifically identified. The following provisions are particularly relevant:

- the right to disconnect and benefit from safeguards for work-life balance in a digital environment,
- being protected against risks and harm to one's health, safety and fundamental rights in interaction with artificial intelligence systems,
- ensuring a safe, secure and fair online environment where fundamental rights are protected, and responsibilities of platforms, especially large players and gatekeepers, are well defined,
- children and young people should be protected and empowered online.

The EU also has provisions on consumer protection, which may be relevant for ensuring product safety. Of relevance in this respect is a fact that in 2020 the European Parliament commissioned a study on

[23] Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment, available online at: https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a .

[24] Council of Europe, Handbook for policy makers on the rights of the child in the digital environment, abailable online at: https://rm.coe.int/publication-it-handbook-for-policy-makers-final-eng/1680a069f8 .

loot boxes.[25] A lot of relevant work is being done in the context of broader efforts of digitalization, developing the digital single market, etc.

### State level

A number of states adopted charters on digital rights which may be of relevance for preventing overuse of the Internet. Some states took action to ban loot boxes in computer games (Japan, China, Belgium, the Netherlands) or treat games with this feature as gambling and therefore have a higher age requirement for them (Germany).

Some states (France, Italy, Slovakia, Canada) have laws providing for a right to disconnect in the context of working conditions.

In Canada, Ontario Human Rights Commission takes an expansive and flexible approach to defining psychiatric disabilities and addictions that are protected by Ontario Human Rights Code. It considers that the Code protects people with mental health disabilities and addictions from discrimination and harassment under the ground of "disability."[26] In the USA, at least one plaintiff has contended that his PUI was a disability entitled to protection under the Americans with Disabilities Act.[27]

## Minimization of the threat – recommendations

### Recommendations for policy makers

Existing policies can be improved in a number of ways.

- First, in the context of the right to health the necessary health care services should provide for those suffering from Internet addiction. Updating the list of diagnoses to reflect the current stage of mental health research into PUI may be needed.
- Second, targeted policies should be developed to address the needs of specific categories of vulnerable individuals. One such category is children. Other categories are to be defined relying on mental health research into various types of activities that may lead to PUI.
- Third, the potential of anti-discrimination law provisions prohibiting discrimination on the ground of disability or on the ground of PUI more specifically should be explored.
- Fourth, consumer law avenue is to be used to enhance protection against products containing potentially addictive features (marking such products accordingly is one option).
- Fifth, the question of whether the right to disconnect should be limited to the context of working conditions needs to be explored. Reliance on this right to ensure online and offline time balance in the process of education for children is one possible line of inquiry.

[25] A Cerulli-Harms *et al*, "Loot Boxes in Online Games and their Effect on Consumers, in Particular Young Consumers", available online at: https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2020)652727 .

[26] Ontario Human Rights Commission, „Policy on Preventing Discrimination Based on Mental Health Disabilities and Addictions", available online at: https://www3.ohrc.on.ca/sites/default/files/Policy%20on%20Preventing%20discrimination%20based%20on%20mental%20health%20disabilities%20and%20addictions_ENGLISH_accessible.pdf

[27] Pacenza v. IBM Corp., No. 04 Civ. 5831 (PGG), 2009 U.S. Dist. LEXIS 29778, at *2–3 (S.D.N.Y. Apr. 1, 2009).

### Recommendation for business actors

To minimise the threat of Internet addiction, business actors should opt to move towards more responsible technology and to either refrain from technological solutions that capture the user's attention or to mark the products containing such features accordingly.

## The culture of cancelation and social engineering[28]

### Description of the threat

The concept "culture of cancellation", "cancel culture" or "call out culture" refers to the act of socially "canceling" a person, a form of involuntary ostracism caused by others' actions, aimed to invalidate the victims' opinions and virtually their existence in all areas of society. Even though those actions could or not take place in the cyberspace, specialized literature usually refers to this phenomenon within the framework of social networks (such as posting – true of false – comments about socially reprehensible conduct, such as child abuse or gender violence). Obviously, cancel culture is a major problem when it is illegitimately used, and in certain circumstances can affect the foundations of democracy, for example spreading false accusations directed against other candidates in an election process.

"Social engineering" refers to a series of techniques through which those who apply them gain access to information, especially of a confidential nature and predominantly from natural persons who would not normally access it through legitimate means, with the free and full consent of these (e.g., access permits), to use them for their own benefit and to the clear detriment of the victims of these procedures. What scaffolds these types of techniques is the long-established fact that users are the weak link in the system.

Social engineering may or may not be linked to the social cancellation of a person since through social engineering techniques, for example, a scam can be carried out. In principle, this scam can contribute to the social cancellation of the victim.

Through social engineering techniques, an action could perfectly be generated towards the social cancellation of someone who has not committed any wrongdoing. This could occur, for example, through the use of artificial intelligence and information from social networks to generate a smear campaign through fake news regarding a certain person who can present himself as a rival to be defeated by a politician in an electoral contest (for example, these kinds of techniques were used during the 2016 election of the United States president, which was discovered due to the Facebook-Cambridge Analytica scandal).

Although social engineering techniques have been used since long ago in other fields (it is worth remembering the famous hoaxes of George Parker – who sold the Eiffel Tower several times – and Carlo Ponzi – with his pyramid scheme – in cyberspace they found a fertile ground and have diversified exponentially. This is due to multiple technological factors, which have generated countless risks that

[28] Report section written by Oscar R. Puccinelli.

were barely imagined until recently. These are accentuated by big data, cloud computing, the Internet of things, artificial intelligence and machine learning in the web 2.0 and the social networks environment. The risks are varied and include both damage to physical integrity and material damage (for example, the theft of money from an account, the frustration of a contract, etc.) and non-material damage (subsequent moral damage, damage to the reputation, the digital annihilation of the victim, etc.).

## Developments in relation to the threat

Social engineering is today a key factor in the electoral political sphere and is used both to obtain private information about a rival candidate or his inner circle that can be used – and even distorted – against him or her, as well as to identify those users of social networks who could be influenced to change their vote in favour of who is paying for the social engineering work.

As noted previously, among many examples of the use of false information with the purpose of manipulating the population or certain sectors of it, the alleged impact of information disseminated through social media and its impact on several recent elections is often mentioned. Examples include the tight results of the 2016 UK referendum on whether or not it should continue to be part of the European Union ("Brexit") and the presidential election held in the United States of America that same year, where Donald Trump was also elected by a narrow margin. The most recent case of the 2018 Italian elections is also often mentioned, but there is no consensus that the impact of fake news on this process was significant.

Cybercriminals' social engineering attacks are characterized by the use of several techniques, that are used against the victim: reciprocity (something is offered and given, asking then for something in return), authority (the perpetrator presents himself as someone important, for example, a law enforcement officer), urgency (a supposed situation is created to press the victim to do something, such as occurred with the WannaCry cryptoworm in 2017, that affected 230,000 computers in 150 countries), sympathy (projecting confidence and optimism), concession (acknowledging a fault, such as having to deliver a card that should have been replaced some time before, and one is directed to a fake page) and preloading (for example, sending general news about some circumstance and then generating an opportunity about it, getting the victim to click on a link that is sent).

The information can be obtained by certain techniques such as flattering the victim (achieving empathy); releasing uncertain information to be corrected by the victim; showing a supposed ignorance about something that the victim controls, obtaining the needed information through the explanation given by the victim), being a sounding board (for example, telling alleged intimacies to the victim so that they tell us theirs) and formulating appropriate questions (closed – yes or no, open or guided). Prefabricated scenarios are often used, whose staging requires research and planning; understanding body language (projecting confidence in the victim, avoiding for example touching hair, nose or ears, crossing arms) and understanding expressions (recourse to positioning, imitation).

Once the information is obtained by social engineering, it could be used as described above, against the victim of the cybercrime, who can be, for example, a politician who is running for elective office, or a voter who may be influenced based on the personal characteristics that were obtained through the social engineering attack.

### Minimization of the threat – recommendations

Minimizing the risks of social engineering attacks, both to counter an improper cancellation and to prevent the use of personal information to illegally influence a person to vote for a particular candidate, requires concerted action by civil society, governments and technology companies.

In Latin America, the 2016 OAS and IDB Cybersecurity report provides data from the Latin American Cybersecurity Observatory according to which most countries in the region are not prepared to face cybercrime threats; less than a quarter have cybersecurity strategies or cybersecurity command and control centres, and similarly, most prosecutors' offices are not trained or have no effective resources to prosecute cybercrimes.[29]

Due to the specificity that characterizes many computer crimes, most countries have created specific criminal types, and thus, for example, in Argentina, through Law 26,388 computer crimes were introduced in the Criminal Code, although not all of them are linked to assumptions of social engineering, beyond the fact that the use of identity theft techniques in computer systems and illegal access to equipment is currently the most fashionable form of computer crime.

Obviously, government efforts must be maximized to provide the population with greater means so that it can defend itself against these attacks.

Finally, at the individual level, and in order to detect a social engineering attack, it is recommended to pay special attention to the perpetrator, who is characterized by conveying control and trust, granting favours and gifts, using humour and presenting reasonable reasons for giving access to information.

If a person discovers that his personal and or professional data has disappeared from networks such as LinkedIn, Facebook, Twitter, Instagram, Google, etc. these companies have specific channels and procedures that are quite useful, and it is recommendable to use them as soon as the attack is discovered.

## Threats connected to international investments[30]

### Description of the threat

In international investment, the number of threats to human rights is increasing. Their diversity is directly proportional to the type of activities that could make certain human rights vulnerable. We have seen how the assertion of a human right actually represents a need for protection against threats from individuals, groups or even public authorities. It is important to regulate specific issues through instruments of international law. And in the case of international investments, as in other situations, human rights are put at risk when there is no law enforcement mechanism nor a functioning judicial mechanism to defend them. There is not enough regulation of treatment standards that have a direct impact on human rights. This observation is based on cases where the worst threats to citizens' human

[29] IDB, "Cybersecurity: Are We Ready in Latin America and the Caribbean?", available online at: https://publications.iadb.org/en/cybersecurity-are-we-ready-latin-america-and-caribbean .

[30] Report section written by Cristina Elena Popa Tache.

rights come from states. Thus, unequivocal regulation through investment treaties would be very useful.

Considering the diversity, expansion and evolution of the field of international investments, it is obvious that threats to human rights multiply, depending on the specifics of the investment. We live in an era where international investment can be: building and launching satellites, building renewable power plants, building and operating blood plasma fractionation, media services and more. At the same time, the way in which information circulates we know is orchestrated by media or social media service providers, and this is just one example of issues that can have implications for human rights in their diversity. In the process of globalisation, a significant footprint belongs to the national system of regulation, adjudication, management or dissemination of information.[31] In the face of these developments, states generally play an active role, preferring the status of policy maker rather than policy taker in the international community.

A ground-breaking agreement signed in June 2020 was the Digital Economy Partnership Agreement (DEPA). It reaffirms "the importance of corporate social responsibility, cultural identity and diversity, environmental protection and preservation, gender equality, indigenous rights, labour rights, inclusive trade, sustainable development and traditional knowledge". We are therefore at the intersection of human rights and technological development. In the same vein, problems arise where human rights and freedoms may conflict with the acceptance and use of digitisation. Similarly, cultural or traditional differences may require more attention from digital service providers. The danger of domination of national digital industries needs to be acknowledged, as issues of national security may overlap here, and the limits of national security and safety are set by domestic regulation and may relate to certain human rights.

## Developments – do performance requirements have the power to incorporate sufficient human rights protection?

The problem concerns the way in which the standards of treatment and legal protection of international investments are or are not well defined, so as to ensure the full respect of human rights.

Performance requirements are the only treaty-regulated standards at the moment that could be used to ensure respect for human rights when discussing communications or new technologies. In general, international investment treatment standards can be used either through direct application or through a process of modifying an international standard to suit national or regional conditions, so that the adoption and adaptation of international treatment standards results in creating equivalent national investment treatment standards that are substantially the same as international standards in their technical content, but may have (i) certain editorial differences and (ii) differences resulting from conflicts in government regulations or specific requirements industry caused by fundamental climatic, geographical, political, sociological, technological or infrastructural factors or the stringency of safety or security requirements that a particular standard authority deems appropriate.[32]

[31] J. Pauwelyn, R.A. Wessel and J. Wouters (eds), *Informal International Lawmaking* (2012); K.E. Davis et al. (eds), *Governance by Indicators: Global Power through Quantiification and Rankings*, Published to Oxford Scholarship Online: September 2012; and J.E. Alvarez, *International Organizations as Law-Makers*, Published to Oxford Scholarship, 2005.

[32] Categorisation available at: https://en.wikipedia.org/wiki/International_standard .

Specific treatment standards generally relate to particular aspects of an investment, such as monetary transfers, expropriation, and the right of the investor in times of war, revolution, or civil strife.[33] Health crisis situations have been assimilated to those of war.

*About performance requirements*. Conceptually, performance requirements are conditions imposed on investors, whereby they are required by host states to meet certain specified objectives with respect to their operations in the territory of the host state[34], being in reality means of selecting foreign investors materialized in measures requiring investors to behave in a certain way or achieve certain results in the host state. Respect for human rights in the category of communications and new technologies has the potential to be a universal performance requirement, without being prohibited. The vulnerable point is that while these performance requirements are considered by investors as influencing how they choose to carry out their investment activities, host states consider it appropriate to ensure that investments make an efficient and maximum contribution to development and are aligned with the national goals and priorities of the host state. For the moment, governments do not have the obligation to impose requirements on international investors, it is an option they can choose to utilise. And, of course, it is subject to contractual negotiations between the host state and the investor.

*Analysis of investment treaties.* Most Bilateral Investment Treaties (BITs) include provisions regarding the transparency of national legislation, performance requirements, entry and stay of foreign personnel, general exceptions and the extension of national and most-favoured-nation treatment for the entry and establishment of investments.

According to UNCTAD statistics, the content of BIT provisions varies considerably, even between BITs signed by the same state, reflecting different approaches and, implicitly, different negotiating positions. Restrictions on performance requirements in investment treaties are moving from a narrow restriction to very broad prohibitions.[35] Over the years, in parallel with the development of practice, some provisions have shown a tendency to become more elaborate. From this perspective, a universal extension of this standard is possible by incorporating the obligation to respect human rights in any investment activity of communications and new technologies.

Some BIT models have been prepared by different states, most models being established at the national level, although there are also cases where these models are established bilaterally or even plurilaterally[36](e.g., regionally), reflecting their positions and expectations regarding international norms and standards in the matter. BITs can also influence domestic law.

---

[33] UNCTAD, Bilateral Investment Treaties 1995 – 2006. Trends in Investment Rulemaking, 2007, p. 28.

[34] United Nations Conference on Trade and Development [UNCTAD], 2003, p. 2, apud Suzy H. Nikièma, *Performance Requirements in Investment Treaties Best Practices Series - December 2014,* Ed. International Institute for Sustainable Development (IISD), p. 4.

[35] An exemplary list of BITs that limit or prohibit PRs includes: India-Kuweit BIT (2001), article 4.4; The Japan-India Comprehensive Economic Partnership Agreement (CEPA 2011), article 89; BIT El Salvador – Peru (1996); BIT Bolivia – Mexic (1995); BIT Dominican Republic – Ecuador (1998); Chile – Mexico Free Trade Agreement (1999).

[36] Conform UNCTAD, există, în acest moment, un singur model de BIT stabilit la nivel bilateral (Belgium-Luxembourg Economic Union Model BIT – 2019) și un număr de patru modele de tratate internaționale plurilaterale privind investițiile (cel mai recent fiind SADC - South African Development Community - Model BIT 2012).material disponibil la: https://investmentpolicy.unctad. org/international-investment-agreements/model-agreements, accesat la data de 03.03.2021.

Investment arbitrations in which courts have found violations of this standard are few and far between, and most of the known cases have been based on NAFTA. Therefore, the emergence of a new, different standard would not be far behind the already existing one for which the addition can be made.

### Minimization of the threat – recommendations

Two solutions stand out: a) creation of a new treatment standard; or b) the explicit expansion of the standard regarding the performance requirements by including in the international instruments universal conditions of respect for human rights in relation to any activity related to communications or new technologies.

A new standard of treatment might be phrased in the following manner: the standard of respect for human rights in investments in communications and new technologies. And the substantive focus should be on social issues and culture, with an emphasis on the lack of knowledge about human rights around the world.[37]

As Resolution A/HRC/RES/20/8 (16 July 2012) put it: "Information is a source that activates the economy, making it possible for people to participate in government activities through public forums and contribute to the decision-making process"[38] . From the point of view of social involvement, education is relevant to the respect of human rights. Looking at the digital market sector, we see different ways in which multinational companies and foundations are making major investments in start-ups in this field, or investing in human rights groups, or in social organisations that are bringing, developing and testing new and most appropriate technologies to comply with human rights. Information technologies (artificial intelligence, big data analytics and large-scale automation) are introducing humanity to a new language and projecting the future. With these new technologies come new and hard-to-predict risks to human rights (e.g., non-discrimination, privacy, children's rights, freedom of expression, access to public services and the right to work).[39]

Considering the scenario where performance requirements are properly formulated and applied, developments can be positive when these requirements become effective tools to maximize the economic, environmental and social benefits (including respect for human rights) of foreign investment.

## Ethical and educational aspects of digital rights[40]

[37] Technology and Human Rights (2019). The social and cultural implications of information & communication technology (ICT) on human rights, humanitarian action, and social change. Available online: technologyandhumanrights.org (accessed on May 2019).

[38] United Nations, (2012c). Human Rights Council Resolution 20/8, Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development, A/HRC/RES/20/8 (16 July 2012) . Available online: https://documents-dds-ny.un.org/doc/resolution/gen/G12/153/25/pdf/G1215325.pdf?OpenElement (accessed on January 2017).

[39] Gajendra Sharma, Implementation of Information and Communication Technology for Human Rights Awareness and Promotion, in HighTech and Innovation Journal, Vol.1, Issue 1, 2020, ed. Ital Publication pp. 33-38.

[40] Report section written by Barbora Baďurová, PhD.

### Description of the threat

In order to ensure respect for human rights, policy guidance and legal regulations need to be supported by ethical education and trainings for various target groups. The problems in the digital era regarding human rights as well as moral and legal rights are connected to, for example, anonymity on one hand and technological possibilities of data collection on the other hand. There are therefore many ways in which human rights may be violated – e.g., hate speech and bullying on internet, misusing of information, problem of authorship etc.

People are often tempted by anonymity, the idea of "no identity", or new identity and that affects their actions which are often different from those that would be done in face-to-face interactions. A question emerges – should we rely only on external regulation – law?

In the digital environment, with the temptation of anonymity, it is important to see the role of ethics because ethics, or morality, is often seen as a type of regulation which takes place even when we are practically "alone" and aware that nobody can see our actions, only our conscience. Among ethical theories we can observe interesting ideas pointing out internal and external regulation and autonomous and heteronomous motivation for moral actions (e.g., Kant, Kohlberg, virtue ethics). Authors often tend to point out that doing good (morally desirable) things for their own sake, because they are good as such, not just because of, for instance avoiding punishment, is the most desirable. The ideally moral person is often described as such that her/his intentions, motivation and actions are moral and directed towards that which is right or good.

Moreover, people cannot just mechanically obey rules as there can be situations when the law is not optimal; then they should be regarded more like guidelines. The problem of motivation also arises in relation to aggressors – if the aggressor, criminal is motivated only by avoiding punishment then she/he is just trying to find ways how not to get caught.

The issues regarding digital human rights are related to the problem of control and self-regulation.

## Developments in relation to the threat

One can observe efforts of various international and national organisations addressing related threats. There are several documents oriented at digital ethics and digital citizenship and its promotion also via education. For example, the European Commission created the Digital Education Action Plan (2021-2027).[41] And the Council of Europe has drafted the Digital Citizenship[42] Education Handbook (2019).[43] The United Nations has developed and published in 2021 a guidance on the rights based and ethical use of digital technologies in relation to public health issues.[44]

[41] Digital Education Action Plan (2021-2027) | European Education Area (europa.eu)

[42] Digital citizensip can be understood as 'the right to participate in society online' (Mossberger, Tolbert, & Mcneal, 2007). (6) (PDF) Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference? (researchgate.net)

[43] Council of Europe, Digital Citizenship Education Handbook, available online at: 168093586f) (coe.int)

[44] UNDP-Guidance-on-the-rights-based-and-ethical-use-of-digital-technologies-in-HIV-and-health-programmes-2-EN.pdf

The problem of human rights in the digital age in general has been reflected on by many academics[45] in a multitude of research institutions.[46] Regarding cyber ethics education we can find texts reflecting for instance on the cases of Japan and USA.[47] Digital ethics has been a research topic for experts from different countries in Europe.[48] There are also several ongoing research projects in the area. For example, an Erasmus+ project Ethics4EU focused on European Values for Ethics in Digital Technology contains proposals for education.[49] And the Erasmus+ project Digital, Responsible Citizenship in a Connected World points out some problems of the digital environment and tries to educate the children in particular.[50] However, the problem of ethics education and individual self-regulation is addressed in a majority of the above-mentioned articles and project only implicitly or indirectly. It is worth mentioning that the Erasmus+ project PLATO'S EU[51] is focusing on the intersection between philosophy and the digital environment, and it will also deal with digital ethics and create teaching materials.

## Minimization of the threat - recommendations

"Human rights are better thought of as both moral rights and legal rights".[52] Ideally, we should support and develop also other motivations for desirable actions not just the motivation not to get caught because of a violation of law.

There is a need to support vulnerable groups and their empowerment by education. Education can help vulnerable groups be aware of their rights, dignity, etc. Digital ethical education and trainings should be oriented not only to children but also adults and not just in the context of formal education but also non-nformal education.

It is important that we should contribute to the promotion of ethical behaviour and human rights in the digital environment also by supporting ethical awareness by trainings and education. This can help cultivate more responsible citizens and also develop fully our moral capacity as human beings. The ideal would be to educate people so they can act virtuously, based on ethical or human rights standards not just for avoiding punishment.

Digital education should focus on normative aspects and not just on being able to use technology.[53] Ethics education should also be included – both in the explicit and "hidden" curriculum. Hakimi, Eynon

[45] E.g., Human Rights for the Digital Age: Journal of Mass Media Ethics: Vol 29, No 1 (tandfonline.com)

[46] E.g. Digital Ethics in Times of Crisis: COVID-19 and Access to Education and Learning Spaces (harvard.edu)

[47] Human Rights and Ethics: Concepts, Methodologies, Tools, and Applications ... - Google Knihy p.1622 Japan

[48] John Paul Gibson, Yael Jacob, Damian Gordon, Dympna O'Sullivan, "Developing an educational brick for digital ethics: A case study-driven approach", available online at: https://hal.archives-ouvertes.fr/hal-03377665/document .

[49] https://www.informatics-europe.org/component/phocadownload/category/10-reports.html?download=151:european-values-for-ethics-in-digital-technology .

[50] DRC – Digital, Responsible Citizenship in a Connected World – Digital, Responsible Citizenship in a Connected World. (digital-citizenship.org)

[51] https://platos-eu.org/

[52] Human Rights | Internet Encyclopedia of Philosophy (utm.edu)

[53] Eg. in relation to the young generation: „School pupils – our Digital Natives – have already been acquainted with Internet technologies from early childhood, mostly in the course of entertainment, communication, or the search for information. For many, this interaction with the new technologies appears to already answer the question of whether we need digital education, because the need seems to be addressed through

and Murphy (2021) have pointed out "the lack of evidence particularly for preschool and school-aged children and the disparate communities working in this domain, and we suggest a more cohesive approach, where the wider learning and educational ecosystem is recognized, explicit engagement with ethical theory is central, and mid- to long-term ethical issues are considered alongside immediate concerns".[54]

It would be interesting and important to also focus on adult education, however the question arises who should be responsible for it – media, community, civil society, government? And what form should it take? Possibilities are combining forms of formal and non-formal education and various forms of nudges. Regulation by law as well as self-regulation supported by expert discussion with stakeholders on regular basis.

usage and the acquisition of new skills, and an apparent problem seems to be solving itself." (Digital Education as the Foundation of Digital Ethics in the Interconnected World - Unleashing Creativity in Work & Life - New Work & Digital Communications - Issues – dotmagazine)

[54] Laura Hakimi, Rebecca Eynon, Victoria A. Murphy, "The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape", available online at: https://journals.sagepub.com/doi/full/10.3102/00346543211020116 .

# Threats connected to technology

**Saulius Stonkus, Skirgailė Žalimienė**
VILNIUS UNIVERSITY

**Tiina Pajuste**
TALLINN UNIVERSITY

**Igor Serotila**
AMERICAN UNIVERSITY OF MOLDOVA

**Konstantinos Kouroupis**
FREDERICK UNIVERSITY

**Maria Biliri**
UNIVERSITY OF ATHENS

## Introduction[55]

Technologies that were once believed to help in ensuring human rights can be weaponized by both state and non-state actors. Technology can be used for the surveillance of citizens and the dissemination of disinformation that has the potential to diminish public reliance and trust in scientific data and knowledge. Artificial intelligence is now used daily in a plethora of different ways, including in decision-making in the public and private sectors (e.g. employment, governmental services, in the financial sector and in the justice system), which can pose significant threats to human rights as research has demonstrated that AI may be biased and not able to produce just outcomes in all occasions. AI can also obscure and decrease responsibility for potential human rights violations as it cannot be easily adapted to the traditional mechanisms for holding wrongdoers accountable.

AI can also be embedded in hardware devices, such as drones, which are considered the future of aviation[56] and are one of the key drivers for creating the Digital European Sky[57]. However, along with the introduction of drones into public life, serious privacy issues arise and the ever-increasing usage of drones, due to the scope of increased aerial surveillance possibilities, requires re-estimating the protection of the right to privacy in order to meet the demands of modern society. In this context drones pose a threat: i) first, as a technology, that 'gives wings' (and does it at low cost) to other technologies (e.g., cameras, sound recorders, GPS, infrared and other sensors, etc.), thus allowing to use them in a completely new environment, i.e., in the air, obtaining new surveillance possibilities; ii)

[55] Introduction written by Saulius Stonkus and Prof. Tiina Pajuste.

[56] See Communication from the Commission to the European Parliament and the Council COM(2014)207 "A new era for aviation – Opening the aviation market to the civil use of remotely piloted aircraft systems in a safe and sustainable manner".

[57] See SESAR Joint Undertaking (2020). Digital European Sky Blueprint. Access through: https://www.sesarju.eu/sites/default/files/documents/digital%20european%20sky%20blueprint.pdf.

second, as a platform (base), that integrates various technologies into one whole, including incorporation of AI technology, thus creating qualitatively new surveillance instruments. Therefore, this part of the report also covers threats to privacy connected to drones, in addition to the risks from employing AI.

## Artificial intelligence and risks for online privacy and security[58]

### Description of the threat

Since we are living in the digital age, AI systems and tools are omnipresent in many areas of private and public life. In healthcare, robotic surgery, cardiac ultrasound tools, clinical diagnosis based on machine-learning systems are some of the elements which demonstrate the rise of AI technology. The use of machine learning models to search medical data and uncover insights to help improve health outcomes and patient experiences is occurring almost on a daily basis. In e-commerce, chatbots are used by a large number of companies providing a big variety of services, as it has been already exposed. Smart cities are also a representative example of AI technology.

The aforementioned references are just indicative illustrations of the ongoing and daily use of Artificial Intelligence in our lives. At EU level, AI is the most or at least among the most representative actions of implementation of the EU digital strategy. It reflects the digitalization which penetrates our society in all sectors of private and public life and puts a clear focus on data, technology, and infrastructure.

In addition, AI plays a major role in shaping Europe's digital future. It constitutes one of the most important actions in order to empower people with a new generation of technologies and create a fair and competitive environment for people and businesses. The EU's approach to artificial intelligence centres on excellence and trust, aiming to boost research and industrial capacity and ensure fundamental rights.

The European approach to artificial intelligence (AI) will help build a resilient Europe for the Digital Decade where people and businesses can enjoy the benefits of AI. It focuses on 2 areas: excellence in AI and trustworthy AI. The European approach to AI will ensure that any AI improvements are based on rules that safeguard the functioning of markets and the public sector, and people's safety and fundamental rights.

However, there are many legal and ethical concerns regarding AI, such as: can AI substitute or replace human factor in public and private life? Is AI always trustworthy and can lead to safe conclusions? What is the nature of AI and in which extent it could be used?

### Developments in relation to the threat

Providing a legal framework for technological tools is not an easy process. It hides many risks since technology always precedes law, therefore this work often corresponds mostly to minimize potential dangers and protect fundamental rights and freedoms. AI is actually a young discipline of about sixty years, which brings together sciences, theories and techniques (including mathematical logic,

[58] Report section written by Konstantinos Kouroupis and Igor Serotila.

statistics, probabilities, computational neurobiology and computer science) and whose goal is to achieve the imitation by a machine of the cognitive abilities of a human being.

Albeit the fact that AI is still surrounded by uncertainties, Council of Europe designed a special Committee on Artificial Intelligence (CAI). In fact, this Committee succeeded the ad hoc Committee on Artificial Intelligence (CAHAI) which fulfilled its mandate from 2019 to 2021. The Committee examined the feasibility and potential elements on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law.

AI technologies are expected to bring a wide array of economic and societal benefits to a wide range of sectors, including environment and health, the public sector, finance and justice. However, since AI encompasses directly the lack of human presence, serious concerns arise regarding the protection of privacy and security. Consequently, it will be attempted an exposition of several relevant questions with regard to the aforementioned issues. In its White Paper on Artificial Intelligence, the European Commission explicitly highlighted that AI must be used in conjunction with human rights legislation, and especially in deference to the protection of privacy and data rights. As the possibilities for monitoring and analysing people's daily habits and actions increase, so do the potential implications on human rights.

Facial recognition is a good example of one area where the competing interests of the benefits of AI and its drawbacks are difficult to judge. facial recognition might have two dimensions: identification and authentication of the person. Identification means that the template of a person's facial image is compared to many other templates stored in a database to find out if his or her image is stored there. Authentication (or verification)— often referred to as one-to-one matching — enables the comparison of two biometric templates, usually assumed to belong to the same individual. Two biometric templates are compared to determine if the person shown on the two images is the same person. Such a procedure is, for example, used at Automated Border Control gates used for border checks at airports.

The EDPS expressed some uncertainty regarding how to use facial recognition technology in a way that is compliant with the GDPR's requirements on data minimisation. As methods of facial recognition are not clear, there are question marks what 'necessary' data is to collect. The EDPS has also said that facial recognition is disputable from ethical point of view. The treatment of the human personality as an 'object' clearly violates fundamental human rights, weakening the value of the individual.

## Minimization of the threat - recommendations

Considering the aforementioned, as well as the fact that draft legislation, in form of the Artificial Intelligence Act is underway, in order to minimize the threat of AI on privacy and security the following recommendations should be considered:

- AI systems should be designed to serve mankind and any creation, development and use of AI systems should fully respect human rights, democracy and the rule of law.
- AI development should follow a human rights by-design approach, meaning developers, manufacturers and service providers should assess and document the possible adverse consequences of AI applications on human rights and fundamental freedoms, and adopt

appropriate risk prevention and mitigation measures from the design phase and during their entire lifecycle
- When the potential risks of AI applications are unknown or uncertain, AI development should be based on the precautionary principle
- AI applications should allow meaningful control by human beings over their effects on individuals and society
- AI developers should critically assess the quality, nature, origin and amount of personal data used, reducing unnecessary, redundant or marginal data during AI development and training phases, and monitoring the model's accuracy as it is fed with new data
- Algorithm vigilance should be adopted in order to promote the accountability of all relevant stakeholders by assessing and documenting the expected impacts on individuals and society in each phase of the AI system lifecycle on a continuous basis, to ensure compliance with human rights, the rule of law and democracy
- AI products and services shall be designed in a manner that ensures the right of individuals not to be subject to a decision significantly affecting them based solely on the automated processing of data, without having their views taken into consideration
- In order to enhance users' trust, AI developers, manufacturers and service providers are encouraged to design their products and services in a manner that safeguards users' freedom of choice over the use of AI, by providing feasible alternatives to AI applications.
- Every individual should have a right to be informed appropriately when she or he is interacting directly with an AI system, providing adequate and easy-to-understand information on the purpose and effects of this system
- Every individual should have a right to obtain, on request, knowledge of the reasoning underlying an AI-based decision process where the results of such process are applied to him or her
- The right to object should be ensured in relation to AI systems based on technologies that influence the opinions and personal development of individuals.
- Policy makers should invest resources in digital literacy and education to increase data subjects' awareness and understanding of AI applications and their effects.
- After the adoption of the Artificial Intelligence Act, responsible authorities should look to empower legal safeguards to be applied to all applications of AI systems used for the purpose of deciding or informing decisions impacting the legal rights and other significant interests of individuals and legal persons

# Automated decision making, including profiling[59]

## Description of the threat

Back in 2006, Clive Humby, a British mathematician and data entrepreneur, was the first to declare that "Data is the new oil". A few years later, in 2017, the Economist wrote "The world's most valuable

[59] Report section written by Maria Biliri, University of Athens.

resource is no longer oil, but data". Today, even the EPRS recognizes that "[d]ata may be the new most valuable asset in the modern economy"[60].

The tremendous progress in the field of algorithms has facilitated the exploitation of data available via an expanding pool of sources. Algorithms have been used in both the public and private sectors for generating/extracting knowledge, assisting informed decision-making and nowadays for automated decision-making.

On a European level, the first definition of "automated decision-making" results from the analysis of the Article 15 of the Data Protection Directive[61] stating that "Member States shall grant the right to every person not to be subject *to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work , creditworthiness, reliability, conduct, etc."*

A slightly different and broader definition of "automated (individual) decision-making" results from the Article 22 of the GDPR[62] that repealed the Data Protection Directive. According to this Article "The data subject shall have the right not to be subject to *a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*". Under the GDPR, "profiling" defined in the Article 4 (4) as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" constitutes a type/form of automated decision-making.

A more precise definition of "automated decision-making" is proposed by ELI[63]: "ADM is a (computational) process, including AI techniques and approaches, that, fed by inputs and data received or collected from the environment, can generate, given a set of pre- defined objectives, outputs in a wide variety of forms (content, ratings, recommendations, decisions, predictions, etc.)". Some examples of automated decision-making include a) the use of an automated face recognition system in a football stadium to prevent the entrance of banned spectators in Denmark, b) profiling job applicants based on their personal emails in Finland, c) allocating treatment for patients in the public health system in Italy, d) detecting welfare fraud in the Netherlands, e) credit scoring systems globally.

Automated decision-making could entail social and economic benefits, including efficiency, reduction of transaction costs, improvement in quality of goods and services, simplification and acceleration of

[60] European Parliamentary Research Service, Briefing "Is data the new oil? Competition issues in the digital economy", 2002, available online: https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/646117/EPRS_BRI(2020)646117_EN.pdf

[61] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[62] 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, and repealing Directive 95/46/EC (General Data Protection Regulation).

[63] European Law Institute, ELI Innovation Paper, Guiding Principles for Automated Decision-Making in the EU, 2022, available online: https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Innovation_Paper_on_Guiding_Principles_for_ADM_in_the_EU.pdf.

procedures implemented by public and private entities, prevention of crime/unlawful acts and effective public administration.

However, it might have negative, even "catastrophic", impacts on human rights, including, but not limited to, the right to privacy and associated rights in key sectors such as "a) law enforcement, national security, criminal justice and border management, b) public services, c) employment, d) (online) content management [64]. Due to the algorithmic opacity (algorithms are often called "black boxes"), the probabilistic nature of the predictions, the risks related to the quality of the data used as inputs and to the determination of the objectives, automated decision-making might result in restricting freedom of choice, undermining human dignity and autonomy, reducing diversity, enlarging bias and discrimination as well as perpetuating stereotypes and social segregation.

## Developments in relation to the threat

As already described, the GDPR provides the right not to be subject to solely automated decision-making for individuals who "are in the [European] Union". In 2018, the Article 29 Data Protection Working Party (now EDPB) issued Guidelines aiming to clarify GDPR provisions that "address the risks arising from profiling and automated decision-making, notably, but not limited to privacy[65]". In these Guidelines, the EDPB explained that "The term right in the provision does not mean that Article 22(1) applies only when actively invoked by the data subject. Article 22(1) establishes a general prohibition for decision-making based solely on automated processing. This prohibition applies whether or not the data subject takes an action regarding the processing of their personal data". Furthermore, some EU Member States have adopted innovative approaches into their GDPR implementing laws laying down supplementary requirements for automated-decision making, namely: a) conducting a human rights impact assessment (Slovenia); b) providing for the right to legibility/explanation about the algorithmic decisions (France, Hungary) and c) ensuring human intervention on algorithmic decisions through an effective accountability mechanism (Ireland).

The EU has adopted/proposed adoption of sector-specific rules concerning certain forms/types of automated decision-making that focus on accountability, transparency and safety - implementation of adequate measures so that human rights risks are eliminated. In particular, the P2B Regulation[66] provides for transparency obligations during the provision of ranking services. Likewise, the proposed DSA[67] – acknowledging risks arising from the algorithmic logic – lays down accountability (including human intervention), transparency and safety obligations relating to recommender systems, terms and conditions and content moderation. The proposed AI Act[68] regulates the development and use of AI systems on a risk basis. Risks arising from AI Systems should be assessed and, then, addressed via

[64] UN Office of the High Commissioner for Human Rights, The right to privacy in the digital age*, 2021, available online:

https://www.ohchr.org › Session48 › Documents

[65] EDPB/WP29, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation

2016/679 (WP251rev.01), as last Revised and Adopted on 6th February 2018 (endorsed by the EDPB).

[66] Regulation (EU) 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.

[67] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

[68] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

the implementation of appropriate measures and compliance with accountability and transparency obligations. As risks increase, stricter rules apply, including the prohibition of placing on the market, putting into service or use of certain AI systems.

The United Nations has touched on automated decision-making and profiling in a report published in 2021 and mandated by the Human Rights Council in its resolution 42/15[69]. In this Report, the United Nations High Commissioner for Human Rights analyses, inter alia, how automated decision-making and profiling affects human rights and provides a set of recommendations for States and businesses so that harmful outcomes are prevented and minimized.

### Minimization of the threat – recommendations

Despite the fact that automated decision-making attracts regulatory attention, relevant rules are partial in their scope, unharmonized and complex. Firstly, a consistent, coherent and all-embracing body of rules on automated decision-making needs to be formed clarifying liability issues. States need to develop cooperation mechanisms towards ensuring that rules governing automated decision-making are properly implemented.

Moreover, Policymakers need to collaborate on creating certification schemes (i.e. seals) and reporting mechanisms to alleviate bias and other relevant problems. In addition, it is crucial that they work together on drawing up Codes of Conduct. Said Codes would, inter alia, contain:

1. draft Human Rights Impact Assessments;
2. methodologies that could be applied to effectively audit decision-making algorithms and facilitate human intervention;
3. draft Notices and relevant guidelines on how to provide individuals with information on the decision-making algorithms using clear and plain language.

Lastly, a great effort should be made to properly educate individuals in order to provide them with the opportunity to familiarise themselves with the algorithmic logic and also raise awareness on human rights issues.

## Aerial surveillance in the digital age: drone related privacy issues[70]

### Description of the threat

Commercial drone[71] industry has flourished in recent years and unmanned aerial vehicle (UAV) technology, which a few decades ago was exclusively a part of modern military equipment, became

[69] UN Office of the High Commissioner for Human Rights, ibid.

[70] Report section written by Assoc. Prof. Skirgailė Žalimienė and PhD student Saulius Stonkus,Vilnius University Faculty of Law, Lithuania.

[71] The term 'drone' is often used to describe virtually any device that is able to fly without a pilot on-board. More technical term is unmanned aerial vehicle (UAV), which in conjunction with the equipment necessary to control it forms unmanned aircraft system (UAS). The latter can be divided into remotely piloted aircraft systems (RPAS) and autonomous aircrafts, which doesn't require any human input during flight at all, whereas RPAS are piloted by remote pilots.

available to the general public. There is no doubt, that drones are very useful. With evolving drone technologies, various new business model opportunities emerge, such as parcel delivery by air, aerial photography, air taxi, drone journalism, etc. Drones offer new services and applications going far beyond traditional aviation and promise an opportunity to perform existing services in a more affordable and environmentally friendly way. Sometimes drones are hard to replace, especially in difficult situations, for example, when carrying out search and rescue missions after natural disasters. The capabilities of drones are almost limitless, making them applicable in any area.

However, modern drones as a rule are equipped with high-end technologies, which can capture and store huge amounts of various data, including private data – not only sound recordings and standard images (like photos and videos), but also thermal images, biometric data, geo-location and geo-spatial data, etc. Whereas various combinations of the aforementioned data enable to create new data in a qualitative sense[72]. Drones can also intercept communications, autonomously track a target, interact with each other and provide such functions as face recognition or identification of vehicle license plate number, even make a live stream, automatically send stored data to other devices (including other drones via 'internet of drones'), upload it directly online[73] and much more. Practice shows that drones can also be hacked and intercepted themselves, including the possibility to retrieve the data stored in drone's internal memory[74]. In addition to this, drones are piloted remotely or in some cases, with advanced artificial intelligence (AI) technology[75], even able to develop their own flight patterns with the only human input being the destination, which makes it very hard to trace the actual drone users[76] and this is crucial in the case of their possible liability. Moreover, modern drones can be as small as the size of a hummingbird and very silent, making them extremely hard to notice. Thus, private data with the use of drones not only can be easily accessed and collected in the areas where people reasonably expect privacy, but it can be done anonymously.

Despite the fact, that some of these technologies used in drones (e.g., cameras, sound recorders, GPS sensors, etc.) are relatively not new, therefore quite well known (including threats, that they pose to privacy), drone technology, due to their ability to fly, remote control and availability to the general public at relatively low price, brings them to the next, it is to say, dangerous level[77]. Along with the mass introduction of drones in our everyday life, sophisticated surveillance techniques emerge, making private data, which is so valuable, more vulnerable than ever. In this sense, as the yellow press with the emergence of instantaneous photographs once was seen as a game changer, requiring to re-

[72] E.g., assigning GPS location data to captured images enables to learn the location of the subject of interest or its movement path, etc.

[73] This in turn raises another issue regarding cross-border exchange of data.

[74] There were numerous reports that military drones, used in war zones, had been hacked and landed by enemy forces. And if state-of-the-art military drones can be hacked, the safety of commercial drones is even more questionable. In fact, since then, several studies showed the vulnerability of drone cyber-security.

[75] AI technology, which is widely used in drone software, by itself requires proper attention, especially in connection with the right to privacy and personal data protection (e.g., due to large amount of data, necessary for 'machine learning', 'black box' related issues, etc.). As far as it concerns AI driven drone technologies, for more details on developments in this regard please see previous Section on AI.

[76] Numerous incidents related to unauthorized drone operations in restricted airspace of nuclear plants in France in last decade have proofed how hard it is to catch the actual pilots of the drones.

[77] Until the thrive of drone technology in 21st century, aerial surveillance with conventional aircrafts was extremely expensive, not as effective and mostly available only to state authorities.

estimate the protection of right to privacy in order to meet the demands of society[78], today drones due to the scope of increased aerial surveillance possibilities invoke such necessity.

In this context drones pose a dual threat to privacy: i) first, as a technology, that 'gives wings' (and does it at low cost) to other technologies (e.g., cameras, sound recorders, GPS, infrared and other sensors, etc.), thus allowing to use them in a completely new environment, i.e., in the air, obtaining new surveillance possibilities; ii) second, as a platform (base), that integrates various technologies into one whole, including incorporation of AI technology, that is often used in drone software, thus creating qualitatively new surveillance instruments. Traditionally the state was seen as the source of these surveillance concerns, but increased usage of modern technologies in public life, including the thrive of commercial drone industry, has created "surveillance capitalism"[79], when private entities control most of the data in these days. Massive deployment of drones into public life, despite all new possibilities and benefits, raises serious privacy and personal data protection issues and may even lead to the so-called 'chilling effect', when individuals feel less free-willing and may perform a form of self-preservation / self-censorship by restricting their behaviour in public (but not only) places when they are or believe that they are being watched[80]. This requires appropriate legal response, especially in the light of increasing public concerns regarding bulk interception.

## Developments in relation to the threat

With drones considered to be the future of aviation[81], there were important regulatory developments in the past few years in order to create a framework for successful drone integration into aviation. Taking the European Union as an example, the new Regulation (EU) 2018/1139 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency (commonly referred to as the "Basic Regulation") was adopted. It brought all aircraft, regardless of their operating mass, into EU competence[82].

Following the adoption of Basic Regulation, the Commission delegated regulation (EU) 2019/945 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems and Commission implementing regulation (EU) 2019/947 on the rules and procedures for the operation of unmanned aircraft were adopted and came into force. Basic Regulation together with the aforementioned two regulations of the European Commission brought in important changes regarding drone operations, especially in relation to the right to privacy and personal data protection. For example, the obligation to register drones and their users and to install direct remote identification systems in unmanned aircrafts was introduced, also courses and exams for remote pilots, which include subjects on right to privacy and data protection, with certificates of the exams valid only for

[78] See Warren, S. D.; Brandeis, L. D. (1890). The Right to Privacy, *Harvard Law Review*, 4(5), 193–220.

[79] See Zuboff, P. S. (2019). The age of surveillance capitalism: the fight for a human future at the new frontier of power. Profile Books, London.

[80] For example, see Clarke, R. (2014). The regulation of civilian drones' impacts on behavioural privacy, *Computer Law & Security Review*, 30(3), 286–305.

[81] For example, see Communication from the Commission to the European Parliament and the Council COM(2014)207 "A new era for aviation – Opening the aviation market to the civil use of remotely piloted aircraft systems in a safe and sustainable manner".

[82] Until then all activities with aircraft lighter than 150 kg (which in the case of drones basically means most of them) were under the regulatory competence of the EU Member States.

limited period of time (currently for 5 years), which means that remote pilots will have to periodically renew their knowledge on this matter, etc.

However, at the same time there are many exceptions from such important mechanisms as registration of drones and their users or direct remote identification of unmanned aircrafts, which does not seem to be well founded and make it quite hard to ensure the effectiveness of these measures while protecting the right to privacy and ensuring personal data protection. For example, according to the Implementing regulation (EU) 2019/947 registration of drone operators and obligation to individually mark unmanned aircraft is mandatory only when operating a drone equipped with a sensor able to capture exclusively *personal* data, which covers only information related to an identified or identifiable natural person ('data subject')[83], leaving aside the data, which strictly doesn't fall within the scope of personal data definition, despite the fact, that collection of such data could infringe one's right to privacy. Moreover, the obligation to register is not applied when using drones, which are considered as toys within the meaning of Directive 2009/48/EC, although the latter also can be fitted with cameras, microphones and various other sensors, available to catch and store both private and personal data. The same goes for direct remote identification system[84], which is essential for effectively dealing with the issue of drone users' anonymity and ensuring their traceability, but the requirement to install it doesn't apply, for example, to drones already made available on the market, despite the fact, that such system can be provided as a separate add-on, which can be retrofitted on drones by their users themselves. Furthermore, the Delegated regulation (EU) 2019/945 sets out the requirements for geo-awareness system, which should alert the remote pilots when a potential breach of airspace limitations is detected so that they can take effective immediate action to prevent that breach in the areas where drone use is restricted, including due to privacy concerns. However, this system is not mandatory, not to talk about geo-fencing system, which is completely left out of EU drone regulations, though it could automatically prevent drones from entering or launching in restricted zones. Drone cyber-security issues are also not covered by these regulations.

Other European (non-EU) countries, such as Norway[85], United Kingdom[86], Iceland[87], have developed their drone regulation in accordance with common EU drone rules. Similar regulations, for example, regarding requirements for drone registration, pilot examination and licensing, built-in remote identification systems and etc., are implemented in United States[88], Canada[89], Australia[90] as well. Generally, developments in legal drone regulation are usually characterized by a greater centralization, covering wider range of UAV's, and by taking a risk-based approach, thus with slight variations in the regulations based on type of unmanned aircraft operations being carried out, i.e., different flight

[83] Article 4 of the General Data Protection Regulation.

[84] 'Direct remote identification' means a system that ensures the local broadcast of information about unmanned aircraft (UA) in operation (including the marking of the UA, operator registration number, geographical position of the UA, the route course and ground speed of the UA, geographical position of the remote pilot or, if not available, the take-off point), so that this information can be obtained without physical access to the UA. The similar technical solution was introduced by US Federal aviation administration in December 2020.

[85] See https://luftfartstilsynet.no/en/drones/.

[86] See https://www.caa.co.uk/consumers/remotely-piloted-aircraft/.

[87] See https://www.icetra.is/aviation/drones/.

[88] See https://www.faa.gov/uas/.

[89] See https://tc.canada.ca/en/aviation/drone-safety.

[90] See https://www.casa.gov.au/knowyourdrone/drone-rules.

purposes (recreational, commercial, etc.), drone size, weight or other specifications (e.g., with or without camera), level of pilot competence, etc. Although there are some differences, these common trends regarding drone regulation worldwide at least when it comes to general rules and principles in relation to drone usage illustrates that there is quite strong consensus on this matter, therefore it is highly plausible that a global (harmonised) framework for drones and their use might be adopted.

### Minimization of the threat – recommendations

Public acceptance is a key to the growth of drone market and in order to achieve it the respect of citizens' fundamental rights, such as the right to privacy and personal data protection, must be guaranteed[91]. For this purpose, harmonized rules, which allow civil drone operations while at the same time guaranteeing the required high level of privacy and personal data protection, must be established at the international level, because potentially differing national approaches to these issues could lead to a significant weakening of the protection of these rights.

However, the aforementioned developments in drone regulations are not sufficient to the detriment of protection of the right to privacy and personal data. Further joint action at the international level should be taken to develop legal requirements, regarding drone and their user's registration, geo-awareness and remote identification systems, also to establish common rules *inter alia* related to drone cyber-security, geo-fencing, etc. Especially when drone manufacturers already apply some of these measures voluntarily (e.g., install geo-fencing systems in their manufactured drones[92]). In addition, more attention should be paid to by design and by default measures (e.g., minimization of data, gathered by drones, automatic anonymization or removal of unnecessary data[93], etc.) and possible obligations for online service providers (e.g., remote signal blocking, restrictions for data sharing, etc.). AI related issues in drone systems and cross-border exchange of data should get a closer look as well.

On the other hand, it must also be borne in mind, that the use of drones may have the opposite effect and, in some cases, even help to promote human rights, as, for example, is the case in drone journalism (e.g., drones can help journalists to document possible human rights violations in war zones or during demonstrations, etc.). Therefore, more restrictions imposed on drone usage in such cases may lead to weakened protection of other human rights, so the right balance must be struck in this regard.

Based on the "control dilemma", elaborated by David Collingridge[94], influencing technological developments is easier when their implications are not yet manifest, but once we know these implications, they are difficult to change. In other words, when a technology is still at an early stage of development, it is possible to influence the direction of its development, but we do not know yet how it will affect society. However, when the technology has become societally embedded, we do know its

[91] See Riga declaration on remotely piloted aircraft (drones) "Framing the future of aviation", Riga, 6 March 2015. Access through: https://eu2015.lv/images/news/2016_03_06_RPAS_Riga_Declaration.pdf.

[92] For example, see https://www.dji.com/newsroom/news/dji-go-app-now-includes-geo-geofencing-system.

[93] In this sense, drone technology should not be regarded only as posing a threat to privacy, but must also be seen as bringing in new possibilities to strengthen its' protection. For example, the AI technology used in drone software can be easily adopted to automatically anonymise private data captured by drone cameras (e.g., people's faces, home addresses, vehicle license plate numbers or GPS coordinates assigned to photos and videos), etc.

[94] See Collingridge, D. (1980). The Social Control of Technology. London: Pinter.

implications, but it is very difficult to influence its development. Therefore, by taking a step-by-step legislative approach, based on current state of technological development of drone systems, having in mind the fast pace of technological progress in comparison to evolution of legal regulation, we risk to lag far behind the technology. As drone usage in modern society keeps growing rapidly, it is of great importance to tackle these new challenges related to the right to privacy and personal data protection in a timely manner and ensure that all the conditions are met for the safe and sustainable emergence of innovative drone services, enabling the industry to thrive and at the same time adequately deal with citizens' concerns.

# Disinformation

**C H Powell**
UNIVERSITY OF CAPE TOWN

**Birgit Schippers**
UNIVERSITY OF STRATHCLYDE

**Irena Barkane**
UNIVERSITY OF LATVIA

**Oscar Puccinelli**
ROSARIO NATIONAL UNIVERSITY

**Jukka Viljanen**
TAMPERE UNIVERSITY

## Disinformation: the concept[95]

Scholars, governments and commentators are using a multiplicity of terms to describe the phenomenon of disinformation. This report uses the term 'disinformation' broadly: we are treating it as an online phenomenon encapsulating the elements of the '*deliberate* creation and sharing of false and/or manipulated information',[96] 'designed, presented and promoted to intentionally cause public harm or for profit'.[97] This conception excludes two related phenomena: these are, first, the spread of false or malicious information offline; and, second, the spread of unintentionally false or inaccurate information,[98] which is better described as misinformation.

However, our use of the term 'disinformation' will also include information which is technically accurate but is shared with malicious intent (also called malinformation, that is, 'genuine information shared with the intention to cause harm).[99] It can therefore include accurate information that instigates violence. A similar approach has been adopted by academics who frame disinformation as 'viral deception', which contains three vectors: manipulative actors, deceptive behaviour and harmful

[95] Report section written by C H Powell, Birgit Schippers, Irena Barkane, Oscar Puccinelli, Jukka Viljanen.

[96] Digital, Culture, Media and Sport Committee (2017–19), Disinformation and 'fake news': Interim Report (HC 363): Government Response to the Committee's Fifth Report https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/163002.htm

[97] T Meyer and C Marsden, Regulating disinformation with artificial intelligence: effects of disinformation initiatives on freedom of expression and media pluralism, European Parliament, 2019 https://data.europa.eu/doi/10.2861/003689

[98] Ibid.

[99] C Wardle and H Derakhshan, Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking (Council of Europe report DGI(2017)09, 2017) 5; https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77.

content.[100] The 'viral deception' approach focuses on the online behaviour rather than the veracity of the content.[101]

While online behaviour is significant and worthy of study in itself, we consider the truth or falsity of the information shared to be a key factor in the harm that disinformation can cause. As we demonstrate below, an important aspect of countering disinformation is the development of mechanisms and processes to determine the veracity of the information available online.

## Threats posed by disinformation

Concerns over disinformation attracted significant attention in the wake of much publicised elections, such as the United Kingdom's Brexit referendum in 2016, the US presidential election of the same year, and the Kenyan election of 2017.[102] The European Union (EU) describes online disinformation practices as 'public harms', specifically harms to the integrity of electoral processes, and 'threats to our way of life,'[103] which undermine trust and confidence in democratic politics.[104] By eroding trust in elected governments, disinformation undermines public programmes that aim to ensure the common good. Current examples which illustrate this point are disinformation practices around Covid, as reported by the EU and monitoring bodies all over the globe.[105]

As we outline below, there are two important points which the Covid cases demonstrate. The first is the enormous difference made by the spreading of information *online* instead of *offline*. The second is that it is not only private individuals, but also governments, which contribute to the spreading of disinformation.

### Disinformation as an online phenomenon

Social media collates stories from multiple sources, changing the focus to the story rather than the source. This practice makes it difficult for people to judge the credibility of information, because 'posts from publications as unlike as the New York Times and a conspiracy site look nearly identical'.[106] Furthermore, traditional gatekeepers are missing as readers choose their material based on

[100] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, "Disinformation and freedom of opinion and expression", A/HRC/47/25, April 13, 2021, https://undocs.org/en/A/HRC/47/25 citing C François, "Actors, behaviors, content: a disinformation ABC" (Transatlantic Working Group, September 2019).

[101] Khan ibid.

[102] C Cadwalladr, 'The great British Brexit robbery: how our democracy was hijacked', *The Guardian* 7 May 2017; Meyer and Marsden (n 2); Wardle and Derakhshan (n 4) 5.

[103] EU Code of Practice on Disinformation, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

[104] EU Code of Practice on Disinformation (n 10); J Bayer, I Katsirea, O Batura, B Holznagel, S Hartmann and K Lubianiec, *The fight against disinformation and the right to freedom of expression*, European Union, 2021 https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf.

[105] European Union, 'Fighting Disinformation', available at https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation_en; Carlos Cortés & Luisa Fernanda Isaza, "The New Normal? Disinformation and Content Control on Social Media during COVID-19", CELE, Palermo University, available at: https://www.palermo.edu/Archivos_content/2021/cele/papers/Disinformation-and-Content-Control.pdf.

[106] Wardle and Derakshan (n 4) 12.

endorsements and social recommendations.[107] Without a ready means of ascertaining the reliability of a source of information, readers rely on friends and family members to guide them through the system.

South Africa's anti-vaccination disinformation is a case in point, where accounts with fewer than 1000 followers authored two thirds of the content containing anti-vaccination hashtags. These small accounts were responsible for 26% more volume in the anti-vaccine conversation than they were in the total vaccine conversation. With just 6% of volume coming from authors with more than 10,000 followers, anti-vaccine conversation appeared to be driven by users with small followings.[108]

## Disinformation spread by governments

It is worth noting that state actors can be equally guilty of spreading disinformation, for example by denying the existence or spread of the disease or by suggesting cures with no proven medical efficacy.[109] Among various examples of disinformation related to the Covid-19 pandemic in the Ibero-American region are the cases of two Argentine national deputies who recommended the use of chlorine dioxide to combat the virus without any evidence that it was effective against the disease and despite the contraindications published by the World Health Organization. Similarly, a Brazilian ministerial body published a video stating that the use of masks was not effective to combat the virus and that it was harmful to health, despite the fact that this was recommended by the World Health Organization. The Presidents of Brazil and Guatemala, and the mayor of Santiago, Chile, also claimed that various treatments with no medical approval were effective or that public transport does not pose a risk of contagion.[110]

## Disinformation and violence

Disinformation can also cause or exacerbate violence, or actively prevent its resolution. An ongoing example is found in the social media (and some state media) coverage of the war in Ukraine.[111] Political violence in the United States (US) and South Africa further illustrates how false information (e.g., that the 2020 US presidential election was stolen) led directly to violence, which threatened democracy directly but also caused loss of life. The attempted insurrection on 6 January 2021 in Washington had been fuelled by the inaccurate claim that Donald Trump had won the US election, and could have overthrown a democratically elected government.[112] The violent riots that erupted in the South

[107] Wardle and Derakshan (n 4) 12, citing Messing, S., & Westwood, S. J. (2014). 'Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online' Communication Research, 41(8), 1042-1063.

[108] Centre for Analytics and Behavioural Change (CABC) *Vaccine Trust Spectrum Report* https://cabc.org.za/wp-content/uploads/2021/10/Vaccine-Trust-Spectrum-Report-Media-Release.docx-4.pdf at 6

[109] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, "Disinformation and freedom of opinion and expression", A/HRC/47/25, April 13, 2021, https://undocs.org/en/A/HRC/47/25.

[110] Center for Studies on Freedom of Expression and Access to Information 'Are public official's lies unsustainable or do they have far reaching effects? A study on the obligations of the State and its officials to prevent the proliferation of disinformation' August 2021, https://www.palermo.edu/Archivos_content/2021/cele/papers/Disinformation-and-public-officials.pdf.

[111] Mart Susi, Wolfgang Benedek, Gregor Fischer-Lessiak, Matthias C. Kettemann, Birgit Schippers, Jukka Viljanen (eds.), Governing Information Flows During War: A Comparative Study of Content Governance and Media Policy Responses After Russia's Attack against Ukraine (Hamburg: Verlag Hans-Bredow-Institut, 2022), GDHRNet Working Paper #4, DOI: https://doi.org/10.21241/ssoar.78580.

[112] See, generally, https://www.aljazeera.com/program/the-listening-post/2022/6/18/the-spectacle-and-scrutiny-of-the-jan-6-hearings.

African provinces of KwaZulu-Natal and Gauteng in the week of 11 July 2021 resulted in over 330 deaths.[113] Violence included damage to vital infrastructure and led to a failure of the rule of law as citizens took up arms and policed their own neighbourhoods while social media fuelled political and racial divides.[114]

## Current countermeasures to disinformation

Since 2017, the United Nations (UN) Special Rapporteur for Freedom of Opinion and Expression, the Representative for Freedom of the Media of the Organization for Security and Cooperation in Europe (OSCE), the Organization of American States (OAS) Special Rapporteur for Freedom of Expression and the Special Rapporteur on Freedom of Expression and Access to Information of the African Commission on Human and Peoples' Rights (CADHP) have collaborated to release joint declarations on disinformation and freedom of expression.[115] In separate endeavours, the European Union produced soft and hard law instruments as well as legislative proposals to address the challenges faced by disinformation.[116]

Current approaches tend to impose different burdens on state officials and private parties. Where private parties are concerned, the focus is often protecting freedom of expression, which needs to be maintained particularly during elections. Government, by contrast, does not have a right to freedom of expression, but a duty first to ensure it does not disseminate false or misleading information itself, second, to respond to and remove disinformation posted by non-state actors when this disinformation can cause a certain level of harm, and, third, to ensure that a single voice does not dominate the marketplace of ideas.

This approach is supported by the jurisprudence of the Inter-American Court, which holds that officials are charged with a more onerous duty to verify the facts, due to their function and the position they occupy in a democratic society. The Inter-American Court ruled in two resounding cases against Venezuela that the exercise of freedom of expression is not the same when it comes to a merely private subject as opposed to public officials, since in a democratic society it is not only legitimate, rather, it is

[113] https://www.aljazeera.com/news/2021/7/22/south-africa-unrest-death-toll-jumps-to-more-than-300.

[114] https://www.bloomberg.com/news/articles/2021-07-20/south-african-economy-set-to-take-3-4-billion-hit-from-riots

[115] https://www.osce.org/fom/302796; https://www.osce.org/files/f/documents/1/e/379351.pdf; https://www.oas.org › basic_documents › declarations; https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclarationDigitalAge_30April2020_EN.pdf; https://www.ohchr.org/en/special-procedures/sr-freedom-of-opinion-and-expression/resources; https://www.ohchr.org/sites/default/files/2022-05/Gender-Joint-Declaration-Freedex.pdf.

[116] See e.g., European Parliament Resolution of 3 May 2018 on media pluralism and media freedom in the European Union (2017/2209(INI)); Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European democracy action plan COM(2020) 790; the European Declaration on Digital Rights and Principles for the Digital Decade COM(2022) 28; the Charter of Fundamental Rights of the European Union (2012/C 326/02); Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016; the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Shaping Europe's digital future' COM(2020) 67; the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions '2030 Digital Compass: the European way for the Digital Decade' COM(2021) 118; Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC; Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act); Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts; Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising.

sometimes the duty of state authorities to rule on matters of public interest. This duty of special care is particularly accentuated in situations of greater social conflict, disturbances of public order or social or political polarization, precisely because of the set of risks that they may imply for certain people or groups at any given time.[117]

In our survey of legislative and policy proposals, we deal first with the normative framework and then the practical measures suggested by these bodies and initiatives.

## The Normative System

### Non-state actors: Rights

Various documents seek to protect the freedom of expression of non-state actors, in particular by discouraging states from using vague or indeterminate criteria when they restrict freedom of expression. The 'Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda' (the Joint Declaration) suggests that states impose restrictions on the right to freedom of expression only if they be provided for by law, serve one of the legitimate interests recognized under international law, and be necessary and proportionate to protect that interest.[118] The Joint Declaration allows restrictions to prevent advocacy of hatred on protected grounds that constitutes incitement to violence, discrimination or hostility (with reference to Article 20(2), International Covenant on Civil and Political Rights). It also protects intermediaries distributing third party content and requires an independent oversight mechanism to rule that third party content should be removed before the intermediaries can be held liable for it. Under the Joint Declaration, state-mandated blocking of entire websites, IP addresses, ports or network protocols are recognised as an extreme measure which can only be justified where provided for by law and where necessary to protect a human right or other legitimate public interest, including in the sense of that it is proportionate, there are no less intrusive alternative measures which would protect the interest and it respects minimum due process guarantees. Finally, governments may not impose content filtering systems which are not controlled by the end users. The 'Joint Declaration on Freedom of Expression and Gender Justice' supports the prohibition of hate speech, applying it specifically to discrimination and sexual and gender-based violence.

The importance of freedom of expression is further supported by the Santa Clara Principles on Transparency and Accountability in Content Moderation, and, in the EU, by the European Parliament's Resolution on Media Pluralism and Media Freedom and the European Charter of Fundamental Rights of the European Union (the Charter).[119] The latter document, the EU's key human rights instrument, provides for the right to privacy (Article 7), protection of personal data (Article 8), and the protection

[117] I/A Court HR. Case of Ríos et al. V. Venezuela. Preliminary Objections, Merits, Reparations and Costs. Judgment of January 28, 2009. Series C n° 194, <https://www.corteidh.or.cr/docs/canes/articulos/seriec_194_esp.pdf>) and Case of Perozo et al. V. Venezuela. Preliminary Objections, Merits, Reparations and Costs. Judgment of January 28, 2009. Series C No. 195, <https://www.corteidh.or.cr/docs/cases/articulos/seriec_195_esp.pdf>.

[118] A position reinforced by The "Joint Declaration on Media Independence and diversity in the digital age", the "Joint Declaration on Politicians and Public Officials and Freedom of Expression".

[119] Charter of Fundamental Rights of the European Union (2012/C 326/02).

of freedom of expression and information (Article 11). The EU Digital Services Act (DSA)[120] updates the liability rules for intermediaries and introduces added due diligence obligations for very large online platforms. Its balanced approach to the liability of intermediaries seeks to establish effective measures for tackling illegal content and societal risks online. The DSA also aims to set a benchmark for a regulatory approach to online intermediaries worldwide if they offer their services in the EU's single market. In return, online intermediaries will benefit from the legal clarity of the liability exemptions and from a single set of rules within the EU.

### Non-state actors: Obligations and Responsibilities

The 'Joint Declaration on Media Independence and diversity in the digital age' states that media outlets and online platforms should enhance their professionalism and social responsibility, e.g., by adopting codes of conduct and fact-checking systems and putting in place self-regulatory systems or participating in existing systems, to enforce them. The 'Joint Declaration on Freedom of Expression and Elections in the Digital Age' states that digital media and online intermediaries should make a reasonable effort to address dis-, mis- and mal-information and election related spam, including through independent fact-checking and other measures, such as advertisement archives, appropriate content moderation and public alerts.

In Latin America, electoral laws and the regulation of political parties, has been established and extend to candidates for public office.[121] These are intended to prevent 'dirty campaigns', including the use of libel and slander, intrusion into a candidate's private life, or inventing 'information'.[122]

The EU's Digital Services Act (DSA) imposes obligations on digital service providers, such as social media or marketplaces, to tackle the spread of illegal content, online disinformation and other societal risks. These requirements are meant to be proportionate to the size and risks that platforms pose to society. Some of these obligations include measures to counter illegal content online and to react quickly, while respecting fundamental rights, including the freedom of expression and data protection. The DSA also compels online platforms to establish a transparency and accountability framework, for example by providing clear information on content moderation or the use of algorithms for recommending content (so-called recommender systems). Further, it bans targeted advertising on online platforms through profiling children or the use of special categories of personal data such as ethnicity, political views or sexual orientation. It also prohibits misleading practices aimed at manipulating users' choices and gives users with the choice to not receive recommendations based on profiling. The DSA has the potential to significantly improve the mechanisms for removing illegal content and disinformation and effectively protecting users' fundamental rights, but there are many implementation challenges ahead.

---

[120] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

[121] Mexico: article 247.2 of the General Law of Electoral Institutions and Procedures; Argentina: article 140 of the National Electoral Code; Honduras: arts. 146 and 148 of the Electoral and Political Organizations Law, in addition to an extraordinary agreement of 2018 of the National Electoral Chamber by which a "Register of social media accounts and official websites of candidates, political groups and highest authorities is created"; Brazil: Article 9 of Resolution 23.610 / 2019 of the Superior Electoral Court and article 323 of the Electoral Code; Peru: Article 42 of Law No. 28094 - Law of Political Organizations.

[122] There were complaints about manoeuvres of this type in the presidential elections in Mexico (2000 and 2006), Colombia (2014) and the Dominican Republic (2015) and in the referendum to modify the Bolivian Constitution regarding presidential re-election (2016), where extensive and unsubstantiated references were made to illegal campaign finance, corruption in the concession of public works or the private life of the candidates.

### State actors: general obligations

The 'Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda' holds that State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda), and should take care to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.

The 'Joint Declaration on Politicians and Public Officials and Freedom of Expression' focuses on the duties of public officials not to disseminate disinformation. Amongst other measures, it encourages states to provide for disciplinary measures to be imposed on public officials who, when acting or perceived to be acting in an official capacity, make, sponsor, encourage or further disseminate statements which they know or should reasonably know to be false and to ensure that public authorities make every effort to disseminate accurate and reliable information, including about their activities and matters of public interest. It further encourages the prohibition of hate speech, that is 'any advocacy of hatred that constitutes incitement to discrimination, hostility or violence'.

The policy documents also emphasise that governments are under an obligation to counteract the spread of disinformation in ways that fall short of direct prohibitions of speech. Thus the 'Twentieth Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade' considers private control as itself a threat to freedom of expression, calling for measures that address the ways in which the advertising-dependent business models of some digital technology companies create an environment which can also be used for viral dissemination of, inter alia, deception, disinformation and hateful expression. It also urges human rights-sensitive solutions to the challenges caused by disinformation, including the growing possibility of 'deep fakes', in publicly accountable and targeted ways, using approaches that meet the international law standards of legality, legitimacy of objective, and necessity and proportionality.

### State actors: elections

The 'Joint Declaration on Freedom of Expression and Elections in the Digital Age' encourages states to ensure that any restrictions on freedom of expression that apply during election periods comply with the international law three-part test requirements of legality, legitimacy of aim and necessity. This entails that there be no prior censorship of the media, administrative blocking of media websites or internet shutdowns. Limits on the right to disseminate electoral statements should conform to international standards, including that public figures should be required to tolerate a higher degree of criticism and scrutiny than ordinary citizens. The media should also be exempted from liability during election periods for disseminating statements made directly by parties or candidates unless the statements have specifically been held to be unlawful by an independent and impartial court or regulatory body, or the statements constitute incitement to violence and the media outlet had a genuine opportunity to prevent their dissemination.

In the EU, the Draft Regulation on the transparency and targeting of political advertising aims to address obstacles to the cross-border provision of online political advertising services in the internal market as well as problems for democratic processes in the context of the internal market.[123]

## Practical measures

It is evident from the general principles set out above that disinformation cannot simply be suppressed or criminalised. This is both to protect human rights and because mere censorship would not be an effective counter to the problem. As we have seen, states can also be guilty of disinformation. When states seek to promote the truth, they may not be trusted because the disinformation already disseminated has turned part of the population against them.

In practical terms, disinformation cannot be countered without an ongoing process of persuasion and ongoing interaction between states and non-state actors. Furthermore, non-state actors themselves need to identify and challenge misinformation. In this regard, information pluralism and the prevention of monopolies enables non-state actors to contest claims made online.[124] Further, because disinformation spreads extremely quickly, it is imperative to address disinformation online before it becomes viral. Without reliable reference points for the validity of online information, there can be no informed citizenry able to make decisions required in a healthy democracy.[125]

Practical steps against disinformation include:

1. Monitoring and fact-checking: carried out by internet communications companies, academia, media, civil society, and independent fact-checking organizations.
2. Investigative responses, which establish the accuracy of online content and provide insights into disinformation campaigns, including its origins, key actors, degree of spread, and affected communities.
3. Curatorial responses, primarily editorial and content policy and 'community standards'.
4. Technical and algorithmic responses, implemented by the social media platforms, video-sharing and search engines themselves, but also through third party tools (e.g. browser plug-ins) or experimental methods from academic research, using algorithms and/or Artificial Intelligence (AI) to detect and limit the spread of disinformation, or provide context or additional information on individual items and posts.
5. De-monetization responses, designed to stop profit and disincentivise the creation of clickbait, counterfeit news sites, and other kinds of for-profit disinformation.

Additional measures, which focus specifically on the targets of the disinformation, include:

[123] Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising

[124] For a thorough explanation of the importance of information pluralism, see NIT S.R.L. v. The Republic of Moldova, NIT S.R.L. v. The Republic of Moldova, 5.4.2022.

[125] Maria D. Molina, S. Shyam Sundar, Thai Le and Dongwon Lee. "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content, *American Behavioral Scientist*, 2021, Vol. 65(2) 180–212, October 14, 2019.

6. Ethical and normative responses carried out at international, regional and local levels involving public condemnation of acts of disinformation or recommendations and resolutions aimed at thwarting these acts and sensitizing the public to the issues.
7. Educational responses which promote citizens' media and information literacy, critical thinking and verification in the context of online information consumption, as well as journalist training.
8. Empowerment and credibility labelling efforts around building content verification tools and web content indicators, in order to empower citizens and journalists to avoid falling prey to online disinformation.[126]

## Concluding suggestions

Any attempt to counter disinformation has to comply with the principles which we have set out in the section regarding the normative system. These include an ongoing and public discussion on the proportionality analysis by which appropriate levels of intervention in the right to freedom of expression and also the right to privacy are determined. While it is important to have a normative framework, such a framework requires effective implementation, and we have set out possible mechanisms in the above section regarding practical measures. We will conclude with five additional observations.

First, while states are an important cog in the anti-disinformation machinery, they find themselves in an uneven power relationship with transnational online media and, in particular, very large online platforms when it comes to demand and control of information. States have a duty under human rights law to protect their citizens from the harm caused by disinformation. But, both due to the power imbalance between themselves and other stakeholders, and because their own credentials are not above suspicion, they cannot act alone.

Second, further actions need to be taken at global level to limit the huge power large technology companies have over people and democracies. The EU DSA, which introduces new mechanisms for removing illegal content and disinformation while seeking to protect users' fundamental rights, including freedom of expression, could serve as a global benchmark for regulatory approaches to online intermediaries at the global level.

Third, private content providers need well trained, well supported and well paid content monitors who scan and delete posts that violate clearly articulated platform guidelines. Such platforms should be required by law to ensure the validity of the information shared on them as quickly as possible. In this regard, we warn against a lenient approach to the intermediaries through whom information is shared.

Fourth, if content providers do not address disinformation, states should work with other stakeholders to identify and counter the disinformation before it can go viral or do significant harm. We recommend the creation and operation of experienced and well-resourced units that respond to disinformation as

[126] These responses are proposed in 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression", Broadband Commission research report on "Freedom of Expression and Addressing Disinformation on the Internet' <https://broadbandcommission.org/publication/balancing-act-countering-digital-disinformation/>.

required. They should identify disinformation and take action to remove and mitigate it. This takes place on the platforms by taking posts down, reporting accounts, requesting bans, as well as sharing and generating accurate information to dispel the disinformation.

Given the extremely harmful potential of disinformation, we propose that criminal prosecution can and should follow where disinformation caused particular harm or posed a particular threat.

Fifth, any meaningful anti-disinformation programme needs to constrain the state as well as private actors. This is both because the state may not always have, or deserve, the trust of the general populace, and also because, even in those cases where state bodies may (currently) be trusted to identify and suppress disinformation in good faith, unaccountable and concentrated power can lead to abuse. The institutional design of the anti-disinfomation therefore needs to include checks and balances to counter the power of both the state and big tech. We suggest that expert units be staffed by people who are not connected to any social media distributers; that is, people with training in the area of disinformation but no financial or political incentive either to suppress or promote particular narratives.[127] These independent units need to monitor and analyse the distribution of information on social media. Such teams should be allocated to specific areas of interest and specialization, including disinformation, incitement to violence, hate speech and xenophobia (broad types of information).

These teams should be located outside of the state, but will need to work in conjunction with organs of state to handle crises and identify emerging threats. Co-operation is necessary because the civil units may not have the capacity to remedy disinformation (which generally requires a real-time response) or to address forms of disinformation which threaten to develop into a crisis such as an insurrection. Government may be needed to provide the person-power and forms of state-sanctioned coercion when strictly necessary. It should should thus have access to the knowledge produced by such bodies, and should have dedicated units gaining it and analysing it to be able to respond in a multisectoral manner (there is, for example, a different set of skills required to combat vaccine hesitancy than there is to prevent posts which are organizing an insurrection).

Co-operation between state and non-state actors should not, in itself, promote abuse of power by the state provided the institutional structure keeps the state in dialogue with other actors and accounting to civil society. A healthy institutional structure will facilitate the interaction of NGOs, the private sector, international organisations, such as the United Nations, and governments. Disinformation is a global phenomenon, and these forms of co-operation allow for a global response.

[127] The South African NGOs 'Centre for Analytics and Behavioural Change' and 'Real411' are good examples of such independent fact-checkers. See https://cabc.org.za/ and https://www.real411.org/.

# Threats connected to privacy and freedom of expression

**Tiina Paiuste**
TALLINN UNIVERSITY

**Nikolas Thomopoulos**
UNIVERSITY OF SURREY

**Pinelopi Troullinou**
TRILATERAL RESEARCH

**Artūrs Kučs**
UNIVERSITY OF LATVIA

**Konstantinos Kouroupis**
FREDERICK UNIVERSITY

**Marijana Mladenoy**
UNIVERSITY BUSINESS ACADEMY IN NOVI SAD

**Igor Serotila**
AMERICAN UNIVERSITY OF MOLDOVA

## Introduction[128]

Two human rights that are particularly vulnerable in the digital context are the right to privacy and the freedom of expression. Firstly, privacy has been eroded in the digital context, often with the assistance of the people themselves. The majority of the people are unaware of the extent to which our privacy is impacted by digital means. Peoples' habits are being tracked; data collected at every click of the mouse. Since most human activities leave behind some kind of digital data trail, it has become increasingly easy to track the behaviour of private individuals.[129] Personal data can reveal very intimate details about a person's character, lifestyle and choices. It is often not realized that personal data can be easily misused or used for commercial purposes. This includes using personal data in ways that were not intended at the time of collection. The economic value of personal data is constantly increasing; thus, personal data has become the cornerstone of various business models both online and offline.[130]

[128] Introduction written by Prof. Tiina Pajuste, Tallinn University.

[129] OECD, "The OECD Privacy Framework" (2013), Chapter 2, Supplementary Explanatory Memorandum to the Revised Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (2013), p 20.

[130] Macenaite, M., "The 'Riskification' of European Data Protection Law through a Two-fold Shift", 8(1) *European Journal of Risk Regulation* 506 (2017), p 506. For more on the commercial value of data, see, e.g., Roessler, B., "Should Personal Data be a Tradable Good? On the Moral Limits

Companies, political campaigns and governments gather individuals' personal data and use analytics to determine past and future patterns.

Secondly, the freedom of expression can be both abused and limited in the online context. Anonymity and the impersonality of the social media and the Internet has resulted in the proliferation of hate speech online. The opportunity to broadcast your point of view to the world in seconds is a mixed blessing. Freedom of expression moving online has also given both governments and (social) media platforms new opportunities to exercise censorship. This part of the report will highlight some of the plethora of risks that arise in relation to the right to privacy and the freedom of expression online.

This part contains three different sections: (a) threats in relation to privacy and data protection in the context of digital mobility; (b) the right to be forgotten as a danger to the freedom of expression; and (c) risks connected to deplatforming.

# Threats in relation to privacy and data protection in the era of digital mobility[131]

## Description of the threat

The right to travel or otherwise the freedom of movement is part of Article 13 of the Universal Declaration of Human Rights (UDHR). Privacy is a protected human right linked with private life worldwide and is contained in Article 8 of the European Convention of Human Rights. However, it varies from a fundamental right (e.g. in India) or constitutional right (e.g. in Israel) to a qualified right in other countries where absolute rights such as national security interfere (e.g. in the United States). In the 21st century, when data has become the global currency[132,133], specific data protection regulations have emerged to protect personal data. The right to privacy has been derived from a broader conceptualisation of data protection legislation. More than 100 countries have introduced the right to protect personal data, which consequently became intertwined with digital human rights. The latter is pertinent in an era of rapid technological advancements in particular within the mobility and transport sector, where digital mobility and transport automation are prevailing.

A range of direct and indirect threats to the data protection and privacy rights have become apparent in what evolves as a global mobility market. The modern way of living further promotes mobility for work, leisure and social relationships. Mobility seems an ever increasing trend as shown in the COVID-19 aftermath, increasing at the same time potential threats to digital rights. Even during the COVID-19 pandemic, when travel was severely restricted, relevant threats emerged regarding privacy and data protection by the use of digital mobility apps such as the Strava app imposed by the Alesund local

of Markets in Privacy", in Roessler, B., and Mokrosinska, D. (eds.), *Social Dimensions of Privacy: Interdisciplinary Perspectives* (Cambridge: CUP, 2015).

[131] Report section written by Dr Nikolas Thomopoulos (Department of Tourism and Transport & Observatory for Human Rights and Major Events, School of Hospitality and Tourism Management, University of Surrey) and Dr Pinelopi Troullinou (Trilateral Research)

[132] Costantini, F., Thomopoulos, N., Steibel, F., Curl, A., Lugano, G. and Kováčiková, T. (2020) Autonomous vehicles in a GDPR era: An international comparison. In *Advances in Transport Policy and Planning*, Vol. 5, pp. 191-213, Oxford: Elsevier – Academic Press.

[133] Thomopoulos, N., Givoni, M., Rietveld, P. (2015) *ICT for transport: Opportunities and threats*. Cheltenham: Edward Elgar Publishing.

authority in Norway.[134] The app was used to monitor physical exercise at local schools and was tracking movement, whilst also collecting personal data without explicit consent by the app users. Other cases of concern include monitoring of employees via CCTV or other tracking technologies such as sensors, which are widely used in the transport and logistics industry. The increase of home deliveries during the COVID-19 pandemic, further increased the need to address relevant threats regarding the rights of privacy and data protection of both workers and customers. An additional concern is the non-explicit disclosure of mobility or mental disabilities to third parties.

Although GDPR[135] and equivalent regulations in other countries (e.g. Brazil) aim at addressing such threats, the multifaceted nature of these threats and their complex technological eco-system requires expert review. For example, digital mobility platforms promoting Mobility-as-a-Service include by default a wide range of transport providers, often based in different countries or even continents, offering taxi, bus, train, bicycle, e-scooter travel options. This creates the problem of which jurisdiction applies regarding the exercise of certain GDPR provisions[136] (e.g., the right to data erasure or the right to be forgotten) when offering mobility services within the EU.[137] Therefore, exercising user rights becomes increasingly complex and expensive due to the jurisdictional challenges mentioned. At the same time certain companies tend to exploit this lack of regulation or enforcement in practice, since it appears to be profitable to compete in a global market where compliance is not required or at least not able to be enforced effectively.

Furthermore, automated and autonomous vehicles (AVs) produce more than 4TB of data daily, including personal data such as locations visited by specific individuals at specific times. The emergence of large databases facilitated through the use of cloud services, alongside the increasing use of Artificial Intelligence (AI), introduces new social and ethical threats, e.g., the creation of a surveillance [138] society in conjunction with cybersecurity risks.[139] Since such emerging large datasets are also being sold to advertising organisations beyond national borders via various platforms, it is apparent that it becomes virtually impossible for individual users to exercise their data protection and privacy rights in full[140,141]. The latter is due to the unavoidable difficulty in defining who is responsible and needs to address such threats. Consequently, enforcement is problematic as certain regulatory requirements are difficult to implement in practice, particularly in the transport sector, e.g., responsibility when transferring mobility service data outside the EU when all cloud services used in, e.g., Norway are located in the US. Technological developments in mobility and transport e.g. drones or AVs illustrate this cross-border challenge vividly.[142]

---

[134] https://www.datatilsynet.no/en/news/2021/alesund-municipality-fined-for-use-of-strava/

[135] The General Data Protection Regulation which was introduced across the European Union in May 2018.

[136] Thomopoulos, N. (2021) Mobility & Digital Human Rights, 29th June 2021, GDHRNet Highlights Lectures.

[137] For more detail, see the section regarding the right to be forgotten, below.

[138] Herzogenrath-Amelung, H., Troullinou, P. and Thomopoulos, N. (2015) Reversing the order: Towards a philosophically informed debate on ICT for transport. In *ICT for Transport*. Edward Elgar Publishing.

[139] For more detail, see the section regarding AI in the part regarding technology, above.

[140] Judin, T. (2022) GDPR: 4 years on, *GDHRNet workshop,* 5-7 September 2022, Oslo.

[141] Zuboff, P. S. (2019) The age of surveillance capitalism: the fight for a human future at the new frontier of power, London: Profile Books.

[142] For more detail, see the section regarding drones in the part regarding technology, above.

## Developments in relation to the threat

The European Union has been working to address some of these emerging threats since it is the largest single market globally, where GDPR and advanced digital mobility options co-exist. The Data Governance Act[143] (DGA) has entered into force since June 2022 and will be applicable from September 2023. Based on the DGA, the European Commission anticipates that it will contribute in *"saving more than 27 million hours of public transport users' time and up to €20 billion a year in labour costs of car drivers thanks to real-time navigation"144.* It is important to stress that: *"Both personal and non-personal data are in scope of the DGA, and wherever personal data is concerned, the General Data Protection Regulation (GDPR) applies"*[8]. Yet, the definition of personal data still varies across countries. The Digital Services Act[145] (DSA) is expected to become effective from 1 January 2024 aiming at protecting consumers and their fundamental rights online. Nonetheless, the fact that it only focuses on consumers limits its scope from the outset. Similarly, the Multimodal Digital Mobility Services[146] (MDMS) Act aims to offer support for corporations and consumers, particularly focusing on digital mobility providers who offer a unified mobility service to end-users in collaboration with other mobility providers. Lastly, the Artificial Intelligence Act (AIA) is also anticipated to affect developments in this field[147].

Existing Human Rights charters are largely non-binding, (e.g. the Charter of Human Rights and Principles for the Internet or the Charter of Digital Fundamental Rights of the EU), despite the fact that certain treaties are based on fundamental rights (e.g. UDHR).[148] The development of certain regulatory frameworks such as the GDPR in the EU have been developed aiming at protecting relevant human rights, with divergent degree of success to date. Similar regulatory frameworks though do not exist in all countries. Nonetheless, as outlined in the report section on the right to be forgotten[149], European legislation based on Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms advocates for a high level of protection of the right to privacy.[150] Equally, the High Level Expert Group on Artificial Intelligence set up by the European Commission has defined privacy as *"a fundamental right, particularly affected by AI systems"*.[151] Privacy Impact Assessments are recommended to be undertaken to assess relevant risks, but this has proven difficult to implement and enforce in practice, particularly for SMEs (Small- and Medium-sized Enterprises) which constitute 99% of the companies within the EU.[152] Given the major challenges faced by SMEs in Europe since 2019

[143] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

[144] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained

[145] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

[146] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13133-Multimodal-digital-mobility-services_en

[147] For more detail, see the section regarding Artificial Intelligence and risks for online privacy and security, above (by Konstantinos Kouroupis and Igor Serotila).

[148] Tosoni, L. (2022) A review of European law developments, *GDHRNet workshop,* 5-7 September 2022, Oslo.

[149] Section regarding the right to be forgotten, below.

[150] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131

[151] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[152] https://www.europarl.europa.eu/factsheets/en/sheet/63/small-and-medium-sized-enterprises

(e.g. COVID-19, energy crisis), it has proven difficult to address data protection and privacy requirements appropriately. As a result, certain human rights may have been increasingly under threat due to the digitalisation of transport and mobility, which may be further enhanced once the AIA is implemented.

## Minimization of the threat – recommendations

Despite this being a constantly evolving context within Europe and worldwide, certain actions can be taken to minimise risks by the threats outlined. Specific recommendations for diverse stakeholders at local, national, and international level are summarised here:

- Enhance data protection and privacy by design to be at the core of new mobility services and facilitate human rights compliance whilst not hindering innovation. This could be achieved through privacy, societal and ethical impact assessments for certain organisations[153].
- Minimise shared responsibility among digital mobility providers and clarify within the newly agreed Acts in Europe (e.g. MDMS, DSA) to improve enforcement. Standardisation of processes and responsibilities would assist to minimise relevant threats.
- Educate and empower individuals and SMEs regarding their data protection and privacy rights, since it is not right to expect them to carry this burden by themselves. Public and private companies should be competent and responsible by default, whilst also regulated by independent authorities to ensure that certain Human Rights obligations are adhered to.
- Ensure that the same human rights linked with the physical products and services should be linked with digital products and services too. This approach should constitute a core principle across countries and transport modes as advocated by the GDHRNet. This would create a 'phygital' balance, which regulatory sandboxes could aid in defining responsibility and liability. Alternatively, the end-user provider should be liable towards end-users as has been the case with transport providers in the 20th century.
- Establish an international cooperation forum to aid in defining the balance between physical and digital rights, whilst avoiding the so called *'rights inflation'* as advocated by GDHRNet. This could be achieved by linking the balance between rights with appropriate ethical principles facilitated through existing approaches (e.g. SUMINI[154]) and guidelines (e.g. Ethics Guidelines for Trustworthy AI: Privacy & Data Governance[155]).
- Harmonise practices and jurisdictions where possible either by centralising agencies (e.g. DPO at national or supra-national level) or by revising relevant aspects of the GDPR.

These recommendations are by no means prescriptive or a panacea. However, they aim at contributing in better ensuring the co-development of human rights for both the physical and the digital

[153] Stahl, B.C. and Wright, D. (2018) Ethics and privacy in AI and big data: Implementing responsible research and innovation, IEEE Security & Privacy, 16(3), pp.26-33.

[154] Thomopoulos, N. and Grant-Muller, S. (2013) Incorporating equity as part of the wider impacts in transport infrastructure assessment: An application of the SUMINI approach. *Transportation*, *40*(2), pp.315-345.

[155] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

dimensions. It is a duty of all stakeholders involved to contribute in offering advanced digital services (e.g. mobility and transport), whilst complying with existing and emerging human rights.

# Deplatforming[156]

## Description of the threat

Deplatforming', or 'de-platforming', refers to the ejection of a user from a specific technology platform by closing their accounts, banning them, or blocking them from using the platform or its services.

The term is explicitly political because it often refers to banning a user from a platform because of the content of their speech and ideas. As the result of deplatforming the speaker is cancelled, interrupted or otherwise unable to address an audience.[157] It is viewed as an unacceptable, unilateral imposition of power by unregulated 'big tech',[158] form of content-based censorship and prior restraint – forbidding wide range of future speech and allowing bi tech's to manipulate public discourse.

The review of accounts and content can be automated or the result of human review, of a combination of both.[159] It is worth nothing that deplatforming may be permanent or temporary. Temporary suspensions and impossibility to access one's account can be considered as deplatforming.

Platforms justify the removal or banning of a user and/or their content based on violations of its terms of service, thereby denying the user access to the community or service that it offers. Deplatforming can and does occur across a range of platforms and can refer to:

- Social media companies, like Facebook, YouTube or Twitter;
- Commerce platforms such as Amazon of the Apple Store;
- Payment platforms, like PayPal or Visa;
- Service platforms, like Spotify or Stitcher;
- Internet infrastructure services like Cloudflare or web hosting.

While deplatforming negatively affects freedom of expression and is an extreme form of content moderation and a form of punishment for violations of acceptable behaviour as determined by the platform's terms or service or community guidelines. At the same time deplatforming is also viewed as important tool in the arsenal of moderation interventions available to platforms.[160] Deplatforming is justified by appealing to a broader discursive strategy that attempts to halt the normaliazation of potentially harmful speech over time.[161] Deplatforming has been used as a response to hate speech,

[156] Report section written by Artūrs Kučs and Konstantinos Kouroupis.

[157] D. D'Orazio. Deplatforming in Theory and Practice: The Ann Coulter Debacle. In E. Macfarlane, eds., *Dilemmas of free expression* (Toronto: University of Toronto Press, 2022), p. 269.

[158] H. Innes & M. Innes. De-platforming disinformation: conspiracy theories and their control, p. 4. Available at: https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1994631.

[159] C. Radsch. Deplatforming/De-platforming. In L. Belli, N. Zingales, Y. Curzi, eds., *Glossary of Platform Law and Policy Terms* (Brazil: FGV, 2021), p. 109.

[160] S. Jhaver, C. Boylston, D. Yang, A. Bruckman. Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW2, Article 381 2021) p. 381:4.

[161] D. D'Orazio. Deplatforming in Theory and Practice: The Ann Coulter Debacle. In E. Macfarlane, eds., *Dilemmas of free expression* (Toronto: University of Toronto Press, 2022), p. 280.

terrorist content, and disinformation/propaganda. For example, the major social media firms have removed hundreds of ISIS accounts since 2015, seeking to reduce the UN-designated terrorist group's reach online, which forced them onto less public and more closed platforms, reducing their visibility and public outreach, but also making it more difficult to monitor their activities. In 2018, Facebook and Instagram deplatformed (Facebook, 2018) the Myanmar (Facebook, 2018) military after it was involved in the genocide of Rohingya, closing hundreds of pages and accounts related to the military and banning several affiliated users and organizations from its services.

## Developments in relation to threat

In recent years, scholars and mainstream news outlets have frequently criticized social media platforms for not doing enough to efficiently moderate their content. In response, platforms have stepped up their content moderation efforts to curb the online spread of hate speech, misinformation and conspiracy theories. For example, Twitter has labelled tweets that violate its guidelines, including those posted by former US President Trump, as offensive. Pinterest has blocked search results for anti-vaccination queries in an effort to fight the spread of vaccine misinformation. Reddit has banned certain toxic communities and quarantined others. Facebook has identified Nazi and white nationalist groups on its platform and shared them with non-profits who help people leave such hate groups.[162]

At the same time deplatforming has been criticised as raising unprecedented threats to democracy and freedom of expression of citizen. James Titcomb: At one level, a private company is perfectly within its rights to terminate a legal contract for services that a user has entered into voluntarily, on terms and conditions specified by the company. The problem is that a handful of big tech companies have amassed oligopolistic control of social media platforms with global reach and impact. For all the awfulness of Trump's communications, it was extraordinary and unprecedented for an incumbent, democratically elected head of state to be blocked from communicating with tens of millions of followers through the world's most popular online services. As Fraser Myers (2021) commented: "If the tech monopolies can deny a platform to the leader of the free world, then they can deny a voice to anyone".[163] Furthermore, the companies have been criticised for doing this in an environment of limited competition and with little transparency, procedural protection or democratic accountability.[164]

For example, a Dutch court recently allowed a citizen-journalism initiative to sue YouTube for removing some of its videos which had been taken down for violating YouTube's rules on Covid-19 disinformation. While the Court did not order the reinstatement of the videos, it did find that aspects of YouTube's disinformation policy went too far, and were "not permitted" under the right to freedom of expression. Importantly, the Dutch Court recognised that YouTube had a "great responsibility" because it is one of the largest online platforms with a worldwide reach and plays a "dominant role" in public debate online.[165]

[162] S. Jhaver, C. Boylston, D. Yang, A. Bruckman. Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW2, Article 381 2021) p. 381:4.

[163] D. Bromell, Regulating Free Speech in A Digital Age (Springer, 2022), p. 139.

[164] Ibid.

[165] Court of Amsterdam 09.09.2020. judgment in case

Digital Services Act will affect the ability of platforms to remove accounts based on supposed violations of their own terms of service. For example, when platforms disable an account based on a violation of their rules, platforms will be required to provide a "clear and specific statement of reasons" for the decision, including the "facts and circumstances relied on intaking the decision," and "explanations as to why the information is considered to be incompatible" with their policy. Most notably, platforms will be required to have "due regard" to the "fundamental rights" of users under the EU Charter of Fundamental Rights, which guarantees freedom of expression.[166]

Indeed, platforms will be required to set up an internal complaint-handling mechanism, giving users the ability to appeal – free of charge – decisions taken by a platform. Crucially, users will have the right to refer disputes over a decision to an independent out-of-court dispute settlement body, and platforms will be bound by the decision.[167]

As a non-State actor, social media platforms like Facebook has the corporate responsibility to respect human rights under the U.N. Guiding Principle on Business and Human Rights[168] (UNGPs), which includes adherence with the International Covenant on Civil and Political Rights[169] (ICCPR). Article 19 of the ICCPR requires the application of the principles of necessity and proportionality to any measure limiting the right to freedom of expression. This would mean imposing the least intrusive yet *necessary* means in regulating expression to achieve a legitimate aim. The aims that are legitimate are themselves narrow, including the protection of national security, public health and morals, public order, and the rights of others. The application of these standards to social media platforms seeking to regulate users' speech, including State actors' speech, has generated robust debate[170], but the U.N. Special Rapporteur on freedom of expression notes that platforms have an arsenal of tools to proportionately address problematic content.[171] De-platforming or permanent account suspension of a user is the most extreme response.[172]

Schmon and Kuczerawy (2021) suggest that "the doctrine of positive obligations and the horizontal effect of the ECHR could support the argument that rules may be necessary to prevent *arbitrary* decisions by platforms to remove content (or ban users)"[173]

Today there is a growing consensus that we need to update Section 230 of the Communications Decency Act. Facebook's Mark Zuckerberg even told Congress that it "may make sense for there to be

C/13/687385 / KG ZA 20-650 CdK/BB; R Fahy, J. Moller, R. Bellanova. Deplatforming Politicians and the Implications for Europe (2021). Available at: https://globaldigitalcultures.org/2021/02/12/deplatforming-politicians-and-the-implications-for-europe/.

[166] Ibid.

[167] Ibid.

[168] UN. Guiding Principles on Business and Human Rights (2011).

[169] UN. International Covenant on Civil and Political Rights (1966). General Assembly resolution 2200A (XXI)

[170] S. Benesch. But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies. *Yale Journal on Regulation*, 2020. Available at: https://www.yalejreg.com/bulletin/but-facebooks-not-a-country-how-to-interpret-human-rights-law-for-social-media-companies/; N. Hakim. Do Not Trust Facebook to Enforce Human Rights, 2021. Available at: https://opiniojuris.org/2021/03/22/do-not-trust-facebook-to-enforce-human-rights/.

[171] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. A/74/486, 2019.

[172] J. Domino. Beyond the Coup in Myanmar: The Other De-Platforming We Should Have Been Talking About (2021). Available at: https://hrp.law.harvard.edu/myanmar/beyond-the-coup-in-myanmar-the-other-de-platforming/.

[173] D. Bromell, Regulating Free Speech in A Digital Age (Springer, 2022), p. 141.

liability for some of the content," and that Facebook "would benefit from clearer guidance from elected officials."[174] Elected officials, on both sides of the aisle, seem to agree: As a candidate, Joe Biden told the *New York Times* that Section 230 should be "revoked, immediately," and Senator Lindsey Graham (R-SC) has said, "Section 230 as it exists today has got to give."[175] In an interview with NPR, the former Congressmen Christopher Cox (R-CA), a co-author of Section 230, has called for rewriting Section 230, because "the original purpose of this law was to help clean up the Internet, not to facilitate people doing bad things."[176]

How might Section 230 CDA be rewritten? Legal scholars have put forward a variety of proposals, almost all of which adopt a carrot-and-stick approach, by tying a platform's safe-harbour protections to its use of reasonable content-moderation policies. A representative example appeared in 2017, in a *Fordham Law Review* article by Danielle Citron and Benjamin Wittes, who argued that Section 230 should be revised with the following (highlighted) changes: "No provider or user of an interactive computer service *that takes reasonable steps to address known unlawful uses of its services that create serious harm to others* shall be treated as the publisher or speaker of any information provided by another information content provider *in any action arising out of the publication of content provided by that information content provider*."[177]

## Minimization of the threat – recommendations

The platforms are the gatekeepers of information and have numerous effective tools in their arsenal to restrict or disseminate certain information. Therefore, it is reasonable that states have imposed primary duties on platforms to eliminate such harmful content as for instance incitement to violence and hatred. However, initiatives at international and national level aimed at ensuring taking down harmful content from platforms, seems to have forgotten about the states positive obligation to ensure freedom of expression. In mediating the public interest and individual rights online a delicate regulatory balance is required.[178] Thus the international and national law should provide guidance for intermediaries on how to achieve it.[179] The international and national law should set the clear rules in which types of cases deplatforming could be considered proportional measure.

Additionally, any decision about deplatforming should be based on transparent rules. Platforms are required to provide a clear and specific statement of reasons for the decision to suspend or block the account.

Last but not least, there should exist procedural guarantees. The internet intermediaries themselves should have a review procedure in place. At the same time the possibilities to contest decisions of

[174] J. Guynn. Donald Trump and Joe Biden vs. Facebook and Twitter: Why Section 230 could get repealed in 2021. Available at: https://eu.usatoday.com/story/tech/2021/01/04/trump-biden-pelosi-section-230-repeal-facebook-twitter-google/4132529001/.

[175] Interview with Joe Biden (2020). Available at: https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html.

[176] A. Selyukh. Section 230: A Key Legal Shield for Facebook, Google Is About To Change. Available at: https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change.

[177] D. K. Citron, B. Witter. The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity. *Fordham Law Review*, Volume 86, Issue 2, Article 3 (2017).

[178] https://europeanlawblog.eu/2020/04/20/is-it-time-for-europe-to-reassess-internet-intermediary-liability-in-light-of-coronavirus-misinformation/

[179] https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf p.2.

online platforms created should complement, yet leave unaffected in all respects, the possibility to seek judicial redress.[180]

[180] Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC recital 44

# Right to be forgotten – a threat to freedom of expression[181]

## Description of the threat

On May 13, 2014, the Court of Justice of the European Union and the Tennessee district court reached different conclusions in cases dealing with the complex issues of privacy and memory in the Digital Age.[182] Despite the fact that the cases involved a different factual background, the same issue was examined – whether we can restrict the harm caused by digital content that would otherwise be available indefinitely. On one side of the Atlantic, the right to be forgotten (hereinafter: RtbF) was acknowledged by the court and on the other, the court held that true information could not be deleted. However, they shared the idea that the concept of the RtbF raises serious concerns with respect to free expression.[183]

The whole concept refers to information that is in the public domain. Therefore, the right of the public to be adequately informed, in particular on matters of public interest, should be considered as an important parameter. For example, "irrelevant" information should not necessarily mean that personal interests prevail over public interest due to the fact that trivial information for one person is of great importance for another one. Who should judge public interest in this case? Moreover, the RtbF allows individuals to request removal of information that is not false or defamatory, but true and originally published in a fully legal manner, because such data are at the moment not flattering to the concerned individual. Does it mean that we should ignore the right of the public to be adequately informed and move legal and completely true information in order to repackage someone's existence online?

Furthermore, RtbF could be used as an instrument of censorship by making it difficult or impossible to search for relevant articles associated with an individual. Since the aforementioned ruling, a number of controversial links to pages have been removed from Google's search results, though these links are not controversial to everyone. Google removed links connecting British individuals to their criminal convictions but not those of Swiss individuals, and a district court in Amsterdam decided that Google did not need to delete the data because "negative publicity as a result of a serious crime in general is accurate permanent relevant information about a person".

In light of the fact that the GDPR sketches only faint boundaries for the RtbF, search engine operators determine which data deletion requests should be granted and which should be denied without any appropriate guidance. Google issued guidelines for implementing the ruling, which instructs interpretations to be made within existing national law. The guidelines provide substantive direction in the form of criteria for the data-protection authorities' handling of the right to be forgotten complaints. In this regard, allowing private businesses to act as both adjudicators and administrators in matters of freedom of expression is a risky combination that may limit freedom of speech since search engine operators are not well equipped to act as both. Why do we expect search engine

[181] Report section written by Marijana Mladenov and Igor Serotila.

[182] Case C-131/12, Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González (Euro. Ct. of Justice May 13, 2014). Clark v. Viacom Int'l, Inc., No. 3:12–0675, 2014 WL 1934028 (M.D. Tenn. May 13, 2014). Jones, M. L. (2016). Ctrl + Z: The Right to Be Forgotten, NYU Press. p.1.

[183] Jones, M. L. (2016). Ctrl + Z: The Right to Be Forgotten, NYU Press. pp.1-3.

operators to be able to strike a balance between RtbF and freedom of expression without any specified standards related to it and under the pressure that the refusal of the request could lead to enormous fines?

## Developtments in relation to the threat

The GDPR, leaning on a technology-agnostic approach and making room for future innovations, did not bind, in article 17, provisions to current trends and state-of-the-art technologies, intentionally leaving it up to provides to enact standards that will help them determine what information should be removed. The lack of guidelines on how such a complex legal mechanism should be implemented within an already sophisticated and interdependent environment has sparked questions about legitimacy, while commercial reputation-enhancing use of the right has attributed to it the title of a concealed form of censorship from critics.

While criticism of "forgetting" has justification, the mechanism itself is not a novel one, i.e. it has been used previously in criminal law. Furthermore, the lack of consensus on the implementation of the right to be forgotten, opens the door for a diverse set of responses countries can adopt in trying to reconcile the conflict between the right to be forgotten and freedom of speech. In this regard, proportionality can be utilised and factor-based balancing favoured; countries can legally enact and take advantage of various models and legal reasoning available within the international human rights protection system.

No jurisdiction can escape from the difficulty of clarifying criteria in order to strike a fair balance among equally important competing values of privacy and the freedom of expression. While the Court of Justice of the European Union is known for its hard stance regarding the right to data protection, its decisions in *Google v CNIL* and *GC and Others* could be seen as lowering that protection. Namely, the Court did not opt for global de-referencing – the only mechanism capable of guaranteeing complete protection of the right to data protection – and considered the processing of sensitive data by a search engine operator lawful until obtaining a request for de-referencing, even without one of the exceptions enshrined in Art. 9(2) GDPR being fulfilled.

The decisions demonstrated yet again how difficult it is to draw lines in the internet and they will have significant implications not only for internet users, but especially for tech companies in and outside the EU, as many aspects of the judgments directly affect their business models. Furthermore, as the Court is a pioneer when it comes to the right to be forgotten, the decision might also indirectly affect the legislation and court decisions in non-EU States.

Expanding jurisprudence is key in order to enshrine RtbF as a practical and utilitarian right. For example, the judgment of the Supreme Court of Japan on January 31, 2017 set up the substantive requirements of injunctive relief for the removal of certain search results.

The Court established that the illegality of the conduct of providing website information including URLs containing articles with private facts of the said person as part of search results in response to search request terms about the person, should be decided by weighing up various circumstances concerning the "legal interest of not having the said facts published" and "reasons to provide information including the said URLs as search results": [1] the nature and content of the said facts, [2] the extent to which private facts of the said person were distributed as a result of the provision of information including the said URLs, and the extent of the damage specifically suffered by the said person, [3] the social

status and influence of the said person, [4] the purpose and significance of the said articles, [5] the social circumstances at the time of the said articles' publication, and the subsequent changes, and [6] the need to mention the said facts in the said articles, etc.; and as a result, if it is "clear" that the "legal interest of not having the said facts published" is overriding, then it should be reasonably interpreted that it is possible to request the search service operator to remove information including the said URLs from search results.

## Minimization of the threat – recommendations

The right to be forgotten can no longer be seen merely as a right to delete information or to preclude its diffusion, as it was originally the case. In a rapidly advancing technological world, one must understand the right in a more multifaceted way. At its core, the right to be forgotten expands and defines itself as an entitlement for individuals to better control their personal data. Just as for the original pre-Internet right to be forgotten, this entitlement finds its justification in the recognition of the right to personal freedom, dignity, and self-realization.

The following recommendations should be considered in relation to RtbF:

- Continuation of the academic and professional efforts to untangle legal intricacies surrounding the RtbF will help strike a balance with freedom of expression, in order to allow a higher legal standard of protection for citizens, effectively integrate forgetting mechanisms into diverse jurisdictions, as well as tackle expression challenges in a human rights-based approach to development;
- Proportionality instruments can help pave the way for states to enshrine forgetting mechanisms into their legal order, while maintaining an adequate level of protection in respect to freedom of expression of their citizens;
- European courts are in an ideal position to act as the highest instance needed to evolve and transform RtbF into a fully-functioning right; on the other hand, legal inertia and gaps are detrimental for current state of play;
- Clarifying criteria for RTBF need to be developed in order to strike a fair balance among equally important competing values such as privacy and the freedom of expression.

# Threats connected to vulnerable groups

**Tiina Pajuste**
TALLINN UNIVERSITY

**Eva Lievens**
GHENT UNIVERSITY, LAW & TECHNOLOGY

**Vesna Crnić-Grotić**
UNIVERSITY OF RIJEKA

**Šejla Maslo Čerkić**
OSCE MISSION TO BOSNIA AND HERZEGOVINA

## Introduction[184]

As the threats described in this report have demonstrated, digital technology and access to the internet, despite their positive transformative potential, come with many potential risks. Although these threaten us all, vulnerable groups tend to be more affected. Many studies have demonstrated that vulnerable groups (such as refugees, migrants, children, older people, people with disabilities, indigenous groups) benefit less from the digital world.[185] For example, ethnicity, sexual identity, gender, religion and age can all exacerbate incidents of online abuse and harassment. This is especially true of people whose identities cover multiple marginalised groups. Negative experiences online can have a severe impact on that person's willingness to use technology and can thereby deprive the person of meaningful access to the internet.

Vulnerable groups may also have fewer digital skills and access to digital education, which makes them more likely to expose themselves to more risks online. It is vital to be aware of the impact of the digital threats to vulnerable groups, in order to address their needs in the digital realm. This part of the report aims to assist in raising awareness of issues regarding vulnerable groups in the online context by addressing the following three topics: (a) the digital divide and its impact, (b) the rights of children in the digital space, and (c) online hate speech, especially in relation to vulnerable groups.

[184] Introduction written by Prof. Tiina Pajuste, Tallinn University.

[185] See, e.g., UN Secretary-General, Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation, 2020, available online at: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement .

# Digital divide – unequal access to the internet and increased inequality

## Description of the threat

Only 63% of the world's people use the Internet. This means that 2.9 billion people do not have access to the Internet and 96% of them live in developing countries.[186] Although the Covid-19 pandemic has increased the use of the Internet (54% used the Internet in 2019), the impact in 2021 was much smaller than in 2020, so the impact of the pandemic is not a sign of a longer trend.

In an increasing number of countries, a large number of important activities have moved to cyberspace (e.g., job search, teleworking, education, communication with public authorities or medical institutions), making it much easier to participate actively in society if you have access to the internet and the necessary digital skills. However, statistics show that many people do not yet have sufficient internet access, which affects their well-being and rights.

The digital divide is the gap between those who have access to modern information and communication technologies (and the skills to take advantage of them) and those who do not. The digital divide exists, for example, between developed and developing countries, urban and rural populations, young and older, educated and less educated people, and men and women. And while access to computers and the Internet continues to grow, the digital divide persists and, in some cases, even widens. The Covid-19 pandemic underlines the urgency of bridging the digital divide. Digital tools have been a lifeline for millions of people and everyone should benefit from them.[187]

Broadly speaking, the digital divide is linked to six different themes:

1. Lack of infrastructure - lack of appropriate systems and facilities to use the Internet. This is primarily a global problem, with scissors predominantly between countries and within larger countries.
2. Economic reasons – in addition to the problems of the country / region, people's own economic opportunities are a separate problem.
3. Lack of skills – the internet may be available, but often there are no skills to navigate it successfully. The development of digital skills is still in its infancy in the world, but the user base is growing.
4. Linguistic accessibility problems – the Internet is mainly in English. Most social media sites operate in either English or the local main language (Russian, Mandarin), which continues to ignore those who do not speak it.

[186] ITU, "Measuring Digital Development: Facts and Figures 2021", available online at: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf .

[187] See, e.g., UNCTAD, "The Covid-19 Crisis: Accentuating the Need to Bridge Digital Divides", 2020, available at: https://unctad.org/system/files/official-document/dtlinf2020d1_en.pdf

5. Special needs – the Internet is an audiovisual medium that requires user activity. More and more attention is being paid to the fact that the deaf, blind and other people with special needs have been left out of the development of the Internet.
6. Restricting access as a goal – because the internet is so important in today's world, network restrictions are common as a punitive or governmental repressive measure. It is possible to shut down the internet altogether or block access to certain topics.

Internet access is a prerequisite for all other possibilities regarding the Internet to work. The negative impact of the digital divide on human rights has been addressed in the legal literature.[188] People who do not have access to the Internet are often socially isolated, as social interaction is increasingly online or mediated by ICT. Freedom of expression and assembly are increasingly being used online, so if you do not have that opportunity, your rights will be restricted. Secondly, the digital divide is having a negative impact on education. The Internet is a rich body of information, the loss of which is a major obstacle to learning. School tasks increasingly involve the use of computers and the Internet, so the lack of access to these resources can seriously affect children's right to education. Third, the number of jobs requiring digital skills is growing rapidly. The digital divide limits access to these jobs and the associated income. Lack of digital skills can ultimately lead to a complete lack of job opportunities, which can infringe on a person's right to work. The digital divide also exacerbates socioeconomic and other vulnerabilities by barring many people from the information necessary to break out of their current living situation.

Statistics show that vulnerable groups have less access and benefit less from the digital world. This magnifies inequalities and leads to an unjust world, which has been recognised internationally on the highest level. For example, the UN Secretary-General has stated that "Digital divides reflect and amplify existing social, cultural and economic inequalities. The gender gap in global Internet use is a stark example – in two out of every three countries, more men use the Internet than women. ... Similar challenges affect migrants, refugees, internally displaced persons, older persons, young people, children, persons with disabilities, rural populations, and indigenous peoples. We must close these gaps through better metrics, data collection, and coordination of initiatives."[189]

## Developments in relation to the threat

Internet access is a global problem and is handled by a wide variety of organizations. The main international actors in the area are the UN and the European Union. One of the first in-depth discussions within the UN framework was the 2011 report by Frank La Rue, the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression and Opinion, which made a

[188] E.g., Sanders, C.K., and Scanlon, E. "The Digital Divide Is a Human Rights Issue: Advancing Social Inclusion Through Social Work Advocacy". 6 *Journal of Human Rights and Social Work* (2021); McIver Jr, W.J., "A Human Rights Perspective on the Digital Divide", in *Community Practice in the Network Society* (Routledge, 2004).

[189] UN Secretary-General, "Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation", 2020, available online: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement, para 26.

number of recommendations to ensure universal access to the Internet.[190] The report emphasized that ensuring universal access to the Internet should be a priority for all countries.

Since 2012, the UN Human Rights Council has regularly updated its resolution on the promotion, protection and enjoyment of human rights on the Internet, which focuses on the digital divide.[191] The 2016 resolution declared that access to the Internet is a human right.[192] However, this resolution did not address the obligation of states to ensure access to the Internet for all. Instead, the resolution emphasizes that governments should not restrict access. Among the UN's 2015 Sustainable Development Goals is Goal 9c: "significantly increase access to information and communications technology and strive to provide universal and affordable access to the Internet in least developed countries by 2020".

The United Nations has also set up various committees and agencies to address the issue of internet access. These include, for example, the UN Commission on Science and Technology for Development, which has the task of providing high-level advice to the UN through analysis and policy recommendations. An important milestone was the UN Secretary-General's Roadmap for Digital Cooperation 2020. It addresses, inter alia, the achievement of universal connectivity, digital inclusion, digital human rights and digital capacity building. In early 2021, the Office of the Secretary-General's Envoy on Technology was established and now plays a major role in coordinating the issue.

The work of the ITU (International Telecommunication Union) is also very relevant. ITU Strategic Plan2020-2023 includes various goals for Internet access to be achieved by 2023: e.g. 65% of households worldwide must have access to the Internet, 70% of the world's population must use the Internet, access to the Internet should be 25% more affordable (compared to 2017), all countries should adopt a digital strategy and the proportion of people with ICT skills should increase by 40%.[193] The United Nations and the ITU have also set up a Broadband Commission to bring the Internet to more people.[194]

At the end of 2021, the UN Envoy on Technology, together with the United Nations Development Program (UNDP) and the ITU, launched the Multi-Stakeholder Network for Digital Capacity Building, which seeks to increase the level of digital capability worldwide, especially in developing countries, by increasing raising awareness of and facilitating access to existing training opportunities.

Various programs and other commissions have been set up within the United Nations, such as the UNESCO Information for All Program (IFAP), launched in 2001, and the GIGA initiative launched by ITU and UNICEF in 2019, which seeks to ensure that every school in the world has access to the Internet.

---

[190] HRC, "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue", 2011, available online: https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf .

[191] UN press release, 2021, available online: https://www.article19.org/resources/un-human-rights-council-adopts-resolution-on-human-rights-on-the-internet/ .

[192] HRC, "The Promotion, Protection and Enjoyment of Human Rights on the Internet", 2016, available online: https://documents-dds-ny.un.org/doc/UNDOC/LTD/G16/131/89/PDF/G1613189.pdf?OpenElement .

[193] ITU, "Connect 2030 – Goals and Targets", available online: https://www.itu.int/en/mediacentre/backgrounders/Pages/connect-2030-agenda.aspx .

[194] Broadband Commission website: https://broadbandcommission.org/ .

### Minimization of the threat – recommendations

ICT is considered to be one of the most potential tools to help reduce inequalities in the world; but there is also a danger that ill-considered or short-sighted action may increase inequalities.

Global efforts need to consistently be focused at providing everyone access to the internet, to ensure that everyone can benefit from the positive aspects that come from the usage of the internet. To achieve that aim, focus also needs to be placed on the development of digital skills as otherwise the potential negative repercussions (e.g. personal data leaked, becoming victims of scams, etc) might outweigh the positives. The UN Roadmap for Digital Cooperation has the ambition aim of ensuring every person safe and affordable access to the Internet by 2030. More specifically the UN will: [195]

1. Support efforts to establish a baseline of digital connectivity that individuals need to access the online space, as well as a definition of "affordability", including universal targets and metrics;
2. Convene a global group of investors and financing experts to consider the development of a financing platform and find other new models for investment in connectivity, in particular, in hard-to-reach and rural areas;
3. Promote new and potentially transformative models to accelerate connectivity, such as the GIGA initiative of ITU and the United Nations Children's Fund;
4. Promote the development of enabling regulatory environments for smaller-scale Internet providers, along with local and regional assessments of connectivity needs;
5. Accelerate discussions on connectivity as part of emergency preparedness, responses and aid, including working through the inter-agency Emergency Telecommunications Cluster.

These initiatives need to be supported by other organisations, states, other stakeholders and by civil society actions.

It is important to realize that the digital divide is a reality between and within countries, and failing to address it means that all other benefits of the Internet are limited or exacerbate societal inequalities.

## Threats and opportunities for children's rights in the digital environment[196]

[195] UN Secretary-General, Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation, 2020, available online: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement .

[196] Report section written by Prof. dr. Eva Lievens (Ghent University, Law & Technology). This report draws in part on the following article: Lievens, E. (2021). Growing up with digital technologies : how the precautionary principle might contribute to addressing potential serious harm to children's rights. *Nordic Journal of Human Rights*, *39*(2), 128–145, https://doi.org/10.1080/18918131.2021.1992951.

## Description of the threats

Digital technologies have a substantial impact on the lives of children and the rights that are attributed to them by the United Nations Convention on the Rights of the Child (UNCRC), regional human rights texts,[197] and many national constitutions.

There is no doubt that the digital environment has enormous potential for the realisation and exercise of children's rights.[198] Digital devices and services provide children with many opportunities[199] to communicate, consume, share, and create content, often across borders, on social media and through mobile apps. On social networks children can express themselves and establish or nurture relationships. Platforms such TikTok, YouTube and Twitch both allow children to watch entertaining and informative videos and to unleash their own creativity to set up their own video channel, something which was unthinkable in the era of traditional broadcasting. Gaming platforms allow children to play, to consume other cultural content, for instance by attending concerts in the gaming environment, and to brush up on their language skills. Educational technologies and learning platforms allow for personalised learning trajectories and provide support for learning difficulties.[200]

Yet, digital technologies also pose threats to children's rights. Such threats relate to illegal and harmful content and activities that children encounter, on social media, in gaming environments, or even the metaverse.[201] In a recent consultation by the European Union, children expressed their concerns regarding content that glorifies and promotes self-harm, suicide, violence, hate speech, sexual harassment, drug taking, risky online challenges, eating disorders and dangerous dieting practices.[202] Other acts that children can fall victim to relate to cyberbullying and other forms of cyberviolence, online distribution of Child Sexual Abuse Material (CSAM), webcam sexual abuse and grooming,[203] or radicalisation.

Moreover, many of the platforms that provide children with these fora to exercise their rights are built on commercial business models that are data- and advertising-driven,[204] and often not designed with children in mind.[205] Children's data is collected extensively, and their behaviour, attention, and

[197] Such as article 24 of the EU Charter of Fundamental Rights (CFREU).

[198] Article 5 UNCRC.

[199] European Commission (2021). Communication EU strategy on the rights of the child, COM(2021) 142 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0142.

[200] UNICEF (2020). Policy guidance on AI for children. https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf.

[201] Madiega, T., Car, P., Niestadt, M. and Van de Pol, L. (2022) Metaverse Opportunities, risks and policy implications, Study European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733557/EPRS_BRI(2022)733557_EN.pdf.

[202] European Commission (2022) Communication A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+), COM(2022)212 final.

[203] European Commission (2020) EU strategy for a more effective fight against child sexual abuse, COM(2020)607 final.

[204] Verdoodt, V., & Lievens, E. (2017). Targeting children with personalised advertising: how to reconcile the (best) interests of children and advertisers. In G. Vermeulen & E. Lievens (Eds.), *Data protection and privacy under pressure: transatlantic tensions, EU surveillance, and big data* (pp. 313–341). Maklu; Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328; van der Hof, S., Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefaard, T. (2020). The child's right to protection against economic exploitation in the digital world. *The International Journal of Children's Rights*, *28*(4), 833–859. https://doi.org/10.1163/15718182-28040003.

[205] European Commission (2022) Communication A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+), COM(2022)212 final.

emotions are increasingly tracked. Inferences are made about who they are, what they feel, how they perform, and what they would like (to buy), when they browse the internet, play with connected toys or online games, or are active on mobile apps and learning platforms. These inferences are often used to profile them for commercial purposes, for instance, to present them with targeted advertising. Serious concerns have been raised with regard to profiling and other forms of automated decision-making that affect children, especially regarding discriminatory effects and the impact on the right to development.[206] There are fears that algorithms and commercial targeting practices shape children's preferences and thoughts from a very young age[207] and hinder self-development[208] free from commercial motives. Moreover, monetisation elements that sometimes resemble gambling (e.g. lootboxes) or encourage overspending are increasingly integrated in online games. Other concerns relate to influencer marketing, which challenges children's advertising literacy skills, and digital labour by children themselves. Children and their parents are expected to make difficult decisions about whether and how to use such commercial platforms, devices, and services that function in complex and opaque manners[209] and whose impact on the rights of the child are difficult to grasp.

## Developments in relation to the threats

Although the issue of protecting children on the internet popped up on the radar of policymakers around mid-1990s, a more holistic approach that focusses on the full range of children's rights that are impacted in the digital environment has only gained traction in recent years. The Council of Europe's Committee of Ministers adopted a Recommendation on Guidelines to respect, protect, and fulfil the rights of the child in the digital environment in 2018. This was the first comprehensive policy document on this issue adopted by a regional human rights organisation. The recommendation states that the digital environment[210] reshapes 'children's lives in many ways, resulting in opportunities for and risks to their well-being and enjoyment of human rights',[211] and provides states with recommendations on how to review their legislation, policies and practices in order to maximise the potential and minimise the risks. Not much later, the United Nations Committee on the Rights of the Child (CRC Committee) decided to develop a General Comment on the rights of the child in relation to the digital environment. In this General Comment No. 25, adopted in 2021, the CRC Committee emphasises that this environment 'affords new opportunities for the realization of children's rights, but also poses the risks

[206] Article 6 UNCRC.

[207] Article 14 UNCRC. See also: European Data Protection Board (2020). Guidelines 8/2020 on the targeting of social media users. https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_202008_onthetargetingofsocialmediausers_en.pdf: "*The potential adverse impact of targeting may be considerably greater where vulnerable categories of individuals are concerned, such as children. Targeting can influence the shaping of children's personal preferences and interests, ultimately affecting their autonomy and their right to development*".

[208] United Nations Special Rapporteur on the right to privacy (2021). Artificial intelligence and privacy, and children's privacy. https://undocs.org/A/HRC/46/37.

[209] van der Hof, S., Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefaard, T. (2020). The child's right to protection against economic exploitation in the digital world. *The International Journal of Children's Rights*, *28*(4), 833–859. https://doi.org/10.1163/15718182-28040003.

[210] This notion is understood as 'encompassing information and communication technologies (ICTs), including the internet, mobile and associated technologies and devices, as well as digital networks, databases, content and services'.

[211] Council of Europe (2018). Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on guidelines to respect, protect and fulfil the rights of the child in the digital environment. https://search.coe.int/cm/Pages/result_details.aspx?ObjectId= 09000016808b79f7.

of their violation or abuse'.[212] Although the General Comment is not binding per se, it provides an important interpretation of the UNCRC in light of digital developments for its 196 parties. Around the same time, the EU published its Strategy on the Rights of the Child, built on six key pillars of which the digital and information society is one.[213] Within this pillar the EU aims to ensure that children can safely navigate the digital environment, and that its opportunities are harnessed. This more general Strategy was followed in 2022 by a new European strategy for a better internet for kids (BIK+).[214] This BIK+ strategy aims 'to complement and support the practical implementation of the existing measures to protect children online, develop children's skills and empower them to safely enjoy and shape their life online'. Also in 2021, the OECD released its Recommendation on Children in the Digital Environment.[215] This Recommendation, which is aimed at governments, is accompanied by 'Guidelines for Digital Service Providers' intended to support these actors to take actions to protect and respect the rights, safety, and interests of children.

## Minimization of the threat – recommendations

There is a consensus that minimising the risks for children's rights in the digital environment is a shared responsibility of all stakeholders. The policymakers that have adopted the documents described above all agree that enhancing the potential and limiting the threats that digital technologies pose to children and their rights can only be achieved if policymakers and legislators, regulatory authorities (such as data protection and consumer protection authorities), industry, civil society and the research community work together.

According to the CRC Committee, states, for instance, should 'ensure that, in all actions regarding the provision, regulation, design, management and use of the digital environment, the best interests of every child is a primary consideration'.[216] This can be put in practice by reviewing, adopting and updating national legislation (for instance, strengthening data protection frameworks) to address the challenges for children's rights, and by conducting Children's Rights Impact Assessments (CRIAs) in the course of that process to ensure that the full range of children's rights is taken into account.[217] Business actors in the digital sector also need to conduct such CRIAs as part of child rights due diligence, in order to identify and remedy negative impact of their activities on children's rights.[218] Accountability of industry for ensuring that the child's best interests prevail is considered to be essential in an environment that is very much private sector-driven. In this context, concepts such as safety-by-design

[212] United Nations Committee on the Rights of the Child (2021). General Comment No. 25 on the rights of the child in the digital environment, https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC/C/GC/25&Lang=en, para 3.

[213] European Commission (2021). EU Strategy on the Rights of the Child, COM/2021/142 final.

[214] European Commission (2022) Communication A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+), COM(2022)212 final.

[215] OECD (2021). Recommendation on Children in the Digital Environment, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389%20.

[216] United Nations Committee on the Rights of the Child (2021). General Comment No. 25 on the rights of the child in the digital environment, https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC/C/GC/25&Lang=en, para 12.

[217] Ibid. para 23.

[218] Ibid. para 38.

and privacy-by-design are often put forward as solutions to ensure that risks are identified and addressed already during the phase in which technologies are conceptualised, designed and offered to children.

More generally, all policy documents also point to the importance of awareness-raising, education and the provision of information about how digital technologies work, aimed at professionals working with children, parents and children themselves. Involving children in the creation of such campaigns, learning material and information is considered essential.

## Hate speech online and the approach of the Council of Europe and the European Union[219]

### Description of the threat

The threat of hate speech has long been recognized by the Council of Europe as an organization with a mandate to protect human rights and fundamental freedoms. Even though there is no universally accepted international legal definition of hate speech, including within the CoE, and what is hateful is often considered disputed or controversial[220], there is still agreement that most serious abuses of freedom of expression that jeopardize democratic values, social stability and peace need to be properly tackled. Such forms of speech do not enjoy the protection under right of freedom of expression, as defined by international human rights documents.

A number of instruments that have been adopted as well as a consistent case law of the European Court of Human Rights witness to this stance and the efforts undertaken by this organization to prevent and possibly punish hate speech. The situation was, however, exacerbated by the unprecedented spread of the use of social media and online hate speech. The Internet, once perceived as the bastion of free speech and enabler of other rights and freedoms, has provided a platform for discrimination, intolerance, bigotry and hatred towards the most vulnerable minority groups – ethnic, sexual, gender, religious, etc. The advantages of online interaction, including anonymity, accessibility, and affordability have also played in favour of those aiming at spreading hatred.

It is difficult to obtain accurate estimates of the extent of hate speech online due to different regulatory and policy approaches by countries, issues with inconsistent monitoring, "particularly in an internet world which is increasingly user-generated, interconnected, and consisting of multiple forms of content. Personal messages and emails are clearly particularly difficult to track". [221]

However, it is evident that global social, political and technological developments, such as the immigration wave in 2015 together with web 2.0 technical development further contributed to the spread of hate speech online, this time mostly against the migrants arriving from the Middle East or Africa. People with a different culture, religion and the colour of their skins were easy targets for many social media users spreading the feeling of threat against "European values".

[219] Report section written by Prof. dr. sc. Vesna Crnić-Grotić and Dr. Šejla Maslo Čerkić.

220 UN Strategy and Plan of Action on Hate Speech 18 June SYNOPSIS.pdf

[221] Three studies about online hate speech and ways to address it, Council of Europe, October 2014, 16809c85ea (coe.int)

Within the Council of Europe framework, the most relevant definition of hate speech is provided by the Additional Protocol to the Convention on Cybercrime and it is concerned only with hate speech which is racist or xenophobic, defining "«racist and xenophobic material» as any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, colour, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors".[222]

The case law of the Strasbourg system has addressed hate speech in cases concerning, among others, white supremacists telling the others – not white – that they should leave the country; negationism and revisionism of the Holocaust, the use of Nazi symbols, intolerance with respect to Roma, aggressive calls against non-Muslim population, and with regard to digital platforms, the Court considered the failure of state to protect the members of the LGBT community against hate speech as violation of the Convention[223] and sharing prohibited content against ethnic minorities[224].

With regard to cyberhate, it has been defined as "the use of violent, aggressive or offensive language, focused on a specific group of people who share a common property, which can be religion, race, gender or sex or political affiliation through the use of Internet and Social Networks, based on a power imbalance, which can be carried out repeatedly, systematically and uncontrollably, through digital media and often motivated by ideologies.".[225]

The European Commission warned that the hate speech as spread online can have a devastating effect on the fabric of social order, "as it potentially not only negatively affects the groups or individuals that it targets; it also negatively impacts those who speak out for freedom, tolerance and non-discrimination in our open societies and has a chilling effect on the democratic discourse on online platforms."[226]

In particular, online hate speech brought in the responsibilities of online platforms as providers of services and enablers. While the comprehensive regulatory framework addressing (illegal hate) speech online within the EU is yet to be enacted through the Digital Services Act (DSA), the European Commission has, through the launch of the Code of Conduct on Countering Illegal Hate Speech Online, introduced an important self-regulatory mechanism to combat the proliferation of racist and xenophobic speech.

[222] Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, Strasbourg, 28.1.2003.

[223] Beizaras and Levickas v. Lithuania (Coe.Int)

[224] Kilin v. Russia (Coe.Int)

[225] Sergio Andrés Castaño-Pulgarín, Natalia Suárez-Betancur, Luz Magnolia Tilano Vega, Harvey Mauricio Herrera López, Internet, social media and online hate speech. Systematic review, Aggression and Violent Behavior, Volume 58, 2021, 101608, ISSN 1359-1789,https://doi.org/10.1016/j.avb.2021.101608 (https://www.sciencedirect.com/science/article/abs/pii/S1359178921000628#ab0010)

[226] European Commission. 2016b. Code of conduct on countering illegal hate speech online.

European Commission document. http://ec.europa.eu/justice/fundamental-rights/files/hate_

speech_code_of_conduct_en.pdf. Accessed August 12, 2017.

## Developments in relation to the threat

In May 2022, the Council of Europe adopted a new Recommendation CM/Rec(2022)16 with a set of principles and guidelines to its member states aimed at preventing and combatting hate speech, both online and offline. The document builds on the existing framework, standards, case law and the monitoring efforts of the Council of Europe concerning hate speech so far.[227] For the purposes of this recommendation, hate speech is understood "as all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as "race",[2] colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation". Member states are called to address the rise in hate speech, especially in the online sphere, through adoption of a comprehensive legal and policy framework. The Council maintained the existing notion that the most serious forms of hate speech should be addressed through criminal legislation, while other less severe expressions are to be tackled through other means, including civil and administrative law. Forms of speech which are not serious enough to be considered in violation of the Convention should nevertheless be addressed through alternative responses.

Concerning online hate speech, specific guidelines are given to states to ensure clear and foreseeable provisions for the effective removal of speech that is prohibited under criminal, civil and administrative law. In following the guidelines, member states should ensure that freedom of expression is protected as guaranteed under Article 10 and in accordance with the requirements of the European Court of Human Rights. As regards the recent case law of the Court and concerning justified restrictions of online speech, in the recent case of *Lilliendahl v. Iceland* (29297/18), the European Court of Human Rights (ECHR) for the first time posed a direct question of whether applicant's comments (about LGBT population) amounted to hate speech. The Court provided a lengthy explanation of the question in the light of the existing Court's case law. Finding that the speech amounted to less grave forms of hate speech, and therefore considered under Article 10, the Court nevertheless sided with the national court, which noted that the applicant's comments were "serious, severely hurtful and prejudicial", and even though there was no direct call or incitement to violence, it fell outside the protections of the ECHR. Concerning criminal liability of online users for comments posted on social media, in the recent case of *Sanchez v France* , the Fifth Section of the Court of Human Rights held that the conviction of a politician for failing to promptly delete unlawful comments (hateful and racist comments directed at Muslims in France) published by third parties on the public wall of his Facebook account did not breach his rights under Article 10 despite his apparent lack of knowledge of the comments.[228] The decision is based on the principles established in *Delfi v. Estonia* that pertain to the liability of intermediaries. However, interventions by third parties claim that these principles are not suitable to be applied in case of individual users on social media.[229] The decision has been accepted for referral to Grand

[227] Existing CoE treaties and other relevant standard-setting instruments, relevant case law of the European Court of Human Rights and the findings and recommendations of the Council of Europe's monitoring bodies, in particular Recommendation Rec(97)20 of the Committee of Ministers to member States on "hate speech", Recommendation Rec(97)21 of the Committee of Ministers to member States on the media and the promotion of a culture of tolerance and General Policy Recommendation No. 15 on combating hate speech of the European Commission against Racism and Intolerance, as well the broader international and European human rights standards.

[228] Case Law, Strasbourg: Sanchez v France, Politician fined for failing to delete Facebook hate speech, no violation of Article 10 – Inforrm's Blog

[229] Media Defence and EFF intervene in Sanchez v France - Media Defence

Chamber in January 2022[230], but some scholars have already assessed it as "underprotection" of freedom of expression.[231] The Court has previously established in *Kilin v. Russia* that criminal sanction of individual users for racist content posted online whose author was not the user and without personal comments to signify the attitude towards the content is justified and proportionate, even if it was made available to a limited audience on a social network.[232]

In addition to the CoE, the efforts taken by the European Union show the necessity to try to give its own contribution to combatting hate speech online. The EU definition of hate speech that is put forth in the Council Framework Decision 2008/913/JHA of 2008 confines hate speech to "all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin" (Council of the European Union 2008), essentially neglecting such characteristics as sex, gender identity and sexual orientation.[233]

The European Commission reacted by concluding the Code of Conduct with several media platforms like Google, Facebook, Microsoft and Twitter prompting them to act against hate speech in 2016.[234] In 2020, the Commission made an assessment of its impact.[235] It concluded that "in summary, the Code of conduct has contributed to achieve quick progress, including in particular on the swift review and removal of hate speech content[236] In October 2021, the results of the 6th periodical evaluation have been published to show "a mixed picture" of the activity of companies. As reported, IT companies reviewed 81% of the notifications within 24 hours and removed an average of 62.5% of flagged content. These results are lower than the average recorded in 2019 and 2020.[237]

Finally, in her 2020 State of the Union speech, Ms Ursula van der Leyen proposed to introduce hate speech and hate crime on the list of EU crimes as a response to their rise in recent years.[238] Based on that the European Commission took the initiative in December 2021, stating that "hate crime and hate speech are going against the fundamental European values set out in Article 2 of Treaty on EU". Pursuant to Article 83(1) of the Treaty on the Functioning of the EU ('TFEU'), the European Parliament and the Council may establish minimum rules on the definition of criminal offences and sanctions in areas of particularly serious crime with a cross-border dimension." Online hate crime is such trans-border crime by definition.

While the Code of Conduct has brought innovative and fairly comprehensive framework that aimed at tackling hate speech in the EU and was endorsed by the biggest global private actors, its self-regulatory

[230] Grand Chamber Panel's decisions - January 2022.pdf

[231] Liability for Facebook-comments: Why the ECtHR underprotected Freedom of Speech – Leuven Blog for Public Law (leuvenpubliclaw.com)

[232] KILIN v. RUSSIA (coe.int)

[233] (Council Framework Decision 2008/913/JHA) Framework Decision on Combatting Racism and Xenophobia through Criminal Law, prohibiting racist and xenophobic hate crime and hate speech and the efforts needed by competent national authorities to investigate and prosecute hate motivated offences, both offline and online

[234] Fn. supra.

[235] Progress on combating hate speech online through the EU Code of conduct 2016-2019.

[236] (28% of content removed in 2016 vs. 72% in 2019; 40% of notices reviewed within 24h in 2016, 89% in 2019)."

[237] EU Code of Conduct against illegal hate speech online: results remain positive but progress slows down - EU monitor

[238] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

nature and the lack of transparency in decision-making remained its biggest shortcomings. To overcome this, the Digital Service Act (DSA), that is awaiting formal adoption and that will be directly applied in the EU member states is hoped to introduce landmark rules to effectively tackle the spread of illegal content online and protect people's fundamental rights in the digital sphere.[239] With regards to combatting online hate speech, the overarching aim is to ensure more responsibility of intermediary services – social media and marketplaces – which will have to take effective measures to better protect users in the digital environment.

## Minimization of the threat – recommendations

The initiative to make (online) hate crime one of EU crimes may bring about uniformity in the European treatment of hate crime. So far, the Member States had a different approach – from a more lenient one in the Western Europe to the stricter approach in the former socialist countries.[240] Criminalization, however, could not be the only approach. It has to be accompanied with more education and training especially of young people on how to be aware of hate crime, how to recognize it and how to fight it online. In addition, self-regulation remains to be one of the options – however, a more comprehensive regulatory framework will be introduced by the DSA.

The consistent policy with respect to examples of hate crime online is also a prerequisite regardless of its author. Technical difficulties in identifying the author are not easy to overcome so attempts to at least minimize the presence of hate crime in some of the controlled media should be maximized, such as comments on articles in media.

The stances of the European Court of Human Rights in the recent decisions concerning online hate speech also show a trend of narrowing of scope of protected speech under Article 10. This development seems to be a reflection of the global social and political context of migrations and the Covid-19 pandemic. More countries introduce more stringent rules on online speech, in efforts to address numerous concerns, relating not only to hate speech, but also complex misinformation narratives and conspiracy theories' effects on public order and safety. These needs may also be reflected in the future decisions of the European Court of Human Rights.

[239] Digital Services Act: agreement for a transparent and safe online environment | News | European Parliament (europa.eu)

[240] Online Hate Speech in the European Union, ed. Stavros Assimakopoulos *et al*.

# Conclusion and recommendations

Despite the plethora of different threats to human rights protection from the digital environment, the trend of digitalisation will continue and impact us even more profoundly in the years to come. Therefore, it is imperative to keep in mind the risks (both highlighted in this report and others) and take action to minimize them to ensure the benefits of digital technology outweigh the negatives. All actors connected to the digital world need to acknowledge the human rights impact of their actions and weigh the potentially conflicting human rights and other considerations in their decision-making process. If the risks are given due consideration, then digital means can be utilised to ensure human rights protection both online and offline.

On the basis of the analysis of the threats contained in the report, the following recommendations are made. More recommendations can be found at the end of each report section.

## Recommentations regarding overarching issues:

- Consider recognizing **social media platforms as human rights duty-bearers.** The fact that social platforms are corporate entities and are in a position to cause enormous harm to human rights supports the need to recognize them as human rights duty-bearers. Since social platforms enjoy great power and no constitutional responsibility, it is high time to challenge the standard public/private division that still dominates constitutional law across the globe.
- **Develop specific** performance requirements (including human rights requirements) for investors **acting in the digital realm**
- **Concerted action by civil society, governments and technology companies** is required to minimize the risks of social engineering attacks both to counter an improper cancellation and to prevent the use of personal information to illegally influence a person to vote for a particular candidate
- Address the threat of Internet addiction by providing the **necessary health care services and taking preventive measures aimed at better online and offline time balance** inter alia by exploring the potential of a right to disconnect
- Promote **ethical behaviour** and human rights in digital environment by means of **formal and non-formal education**

## Recommendations regarding threats connected to technology:

- **Harmonized general rules regarding technological development must be established at the international level,** because potentially differing national approaches could lead to a significant weakening of the protection of fundamental human rights
- A step-by-step legislative approach, based on the current state of technological development, having in mind the fast pace of technological progress in comparison to evolution of legal regulation, risks lagging far behind the technology, therefore it is important to **develop further legal requirements in relation to the use of AI and drones, focusing more on by-design / by-default measures and possible obligations for online service providers**

- As not all potential risks of AI and drone applications are known or certain, **technological development (and legal regulation) should be based on the precautionary principle**
- **AI and drones** should be designed to serve mankind and effectively **employed to help promote fundamental human rights**
- A consistent, coherent and all-embracing body of **rules on automated decision-making needs to be formed clarifying liability issues**. States need to develop cooperation mechanisms towards ensuring that rules governing automated decision-making are properly implemented
- Policymakers need to collaborate on **creating certification schemes (i.e. seals) and reporting mechanisms** to alleviate bias and other relevant problems. In addition, it is crucial that they work together on drawing up Codes of Conduct

### Recommendations regarding disinformation:

- States have a duty under human rights law to protect their citizens from the harm caused by disinformation. In order to perform that duty, they need to **act together with other stakeholders such as transnational online media a**nd, in particular, very large online platforms
- Action needs to be taken at global level to **limit the huge power large technology companies have over people and democracies.** The EU DSA could serve as a global benchmark for regulatory approaches to online intermediaries at the global level
- Private content providers need well trained and supported content monitors in charge of scanning and deleting posts that violate clearly articulated platform guidelines. Such platforms should be required by law to ensure the validity of the information shared on them
- If content providers do not address disinformation, states should work with other stakeholders to **identify and counter the disinformation before it can go viral or do significant harm**. This could be done through the creation of well-resourced units that would identify and respond to disinformation as required
- Given the extremely harmful potential of disinformation, criminal prosecution should be considered where disinformation has caused particular harm or posed a particular threat
- Anti-disinformation programmes need to constrain the state as well as private actors. The institutional design of the anti-disinfomation therefore needs to include **checks and balances to counter the power of both the state and big tech.** We recommend establishing expert units to monitor and analyse the distribution of information on social media. These teams should not be governmental units, but will need to work in conjunction with organs of state to handle crises and identify emerging threats

### Recommendations regarding threats connected to privacy and freedom of expression:

- Data protection and privacy by design need to be at the core of new mobility services and facilitate human rights compliance whilst not hindering innovation
- **Minimise shared responsibility among digital mobility providers** and clarify within the newly agreed Acts in Europe (e.g. MDMS, DSA) to improve enforcement.

- Ensure that the same **human rights which exist in the physical products and services exist for digital products and services**, which should constitute a core principle across countries and transport modes as advocated by the GDHRNet
- A delicate **regulatory balance is required in international and national law between mediating the public interest and individual rights online**. Law needs to set clear rules in which types of cases deplatforming could be considered a proportional measure
- Decisions about deplatforming should be based on transparent rules and platforms should be required to provide a clear and specific **statement of reasons for the decision to suspend or block the account**
- There need to be **procedural guarantees** – internet intermediaries should have a review procedure in place and there should be a possibility to seek judicial redress
- **Clarifying criteria for RTBF need to be developed** in order to strike a fair balance among equally important competing values such as privacy and the freedom of expression

## Recommendations regarding vulnerable groups:

- **Addressing the global digital divide is most appropriate within the framework of the United Nations** (ITU is particularly important in this field), as the United Nations deals with the topic most comprehensively; regional problems (e.g. related to internet speed and digital skills) should be dealt with in regional institutions
- Digital skills development and training should be prioritised, with **special focus on educating vulnerable groups**
- Efforts aimed at **regulating online speech should at least aim for regional, if not global cooperation**, since threats to vulnerable categories are usually cross-border in terms of their scope and effect
- Policymakers and legislators, regulatory authorities (such as data protection and consumer protection authorities), industry, civil society and the research community must **work together to maximise the potential and minimise the threats that digital technologies pose to children and their rights**
- **States should review, adopt and update national legislation to address the challenges for children's rights, and conduct Children's Rights Impact Assessments (CRIAs)** in the course of that process to ensure that the full range of children's rights is taken into account in the digital environment
- **Business actors in the digital sector also need to conduct CRIAs as part of child rights due diligence**, in order to identify and remedy the negative impact of their activities on children's rights
- To realise children's rights in the digital environment, **awareness-raising, education and information about how digital technologies work should be provided** to professionals working with children, parents and children themselves. **Children should be involved** in the creation of such campaigns, learning material and information

# EU COST Action – CA19143: Global Digital Human Rights Network

The GDHRNet COST Action will systematically explore the theoretical and practical challenges posed by the online context to the protection of human rights. The network will address whether international human rights law is sufficiently detailed to enable governments and private online companies to understand their respective obligations vis-à-vis human rights protection online. It will evaluate how national governments have responded to the task of providing a regulatory framework for online companies and how these companies have transposed the obligation to protect human rights and combat hate speech online into their community standards. The matters of transparency and accountability will be explored, through the lens of corporate social responsibility.

The Action will propose a comprehensive system of human rights protection online, in the form of recommendations of the content assessment obligation by online companies, directed to the companies themselves, European and international policy organs, governments and the general public. The Action will also develop a model which minimises the risk of arbitrary assessment of online content and instead solidifies standards which are used during content assessment; and maximises the transparency of the outcome.

The Action will achieve scientific breakthroughs (a) by means of a quantitative and qualitative assessment of whether private Internet companies' provide comparable protection of human rights online in comparison with judicial institutions, and (b) in the form of a novel holistic theoretical approach to the potential role of artificial intelligence in protecting human rights online, and (c) by providing policy suggestions for private balancing of fundamental rights online.

**Contact:** Dr Mart Susi, Action Chair, mart.susi@tlu.ee

Dr Vygantė Milašiūtė, Action Vice Chair, vygante.milasiute@tf.vu.lt

Mr Gregor Fischer-Lessiak, Science Communications Manager, gregor.fischer@uni-graz