Freedom. Internet. A new opening?

The events of 6 January on Capitol Hill in Washington were an infamous summation of the Trump era and one of the signs of its end. It turned out, however, that they additionally sparked a new discussion - on a global scale - about Internet freedom and freedom of expression online.

The decisions of Twitter and Facebook management boards to block the US president's accounts have provoked very mixed reactions.

Some people asked why it was so late, when during his term in office, Trump posted over 30,000 false messages on his accounts, not to mention hateful and demeaning statements, often referred to as "hurtful speech". His questioning of the election results, with the slogan "the election was stolen", and his calls for a revolt and attack on the Capitol, although not explicitly formulated, were spread online even 2-3 hours before the events. Trump had the support of many organisations such as QAnon, with its leader, the buffalo-clad shaman, Jake Angeli, Proud Boys, America First Policies and Women for America First, with Kylie Jane Kremer, who on 2 January, while promoting the action on Capitol Hill, heated up emotions and wrote on Twitter: "Be a part of History", which Trump sent further with the comment: "I'll be there! Historic day." This continued throughout the 77 days between the election and the decisions by the Congress and the Senate to recognise the election results, as well as the day Biden was sworn in [1]. Undoubtedly, Trump's behaviour threatened the democratic order on a grand scale. And that is why internet companies, citing their codes of ethics and actions, carried out this incredible restriction on the freedom of expression of the president being still in office.

Obviously, it would be better, if other mechanisms and tools made it possible to limit the spread of content dangerous to the social order. Against these decisions, in the name of the aversion to arbitrariness and excessive power in the hands of technology companies, protests have been held all over the world - by civil society organisations, by US business people who believe that Zuckerberg has 'blood on his hands' anyway, and even by the head of the European Commission, Ursula von der Leyen, and the German Chancellor, Angela Merkel. The European Union immediately resorted to language and rhetoric familiar from the numerous protests against GAFA and large US corporations. I am only partly surprised by this. The power of the owners of the large Internet platforms and of the platforms themselves is so enormous that many of their actions escape democratic supervision.

At the same time, the essence of the problem is more complicated than political puffery can show. It is, therefore, important to see the context of this debate, which is relevant to the discussion of the legal framework creating requirements for Internet platforms. They should take responsibility for the content they host and thus face the problem of freedom of expression online without mechanically limiting the great value that user-generated content brings to us all [2].

The reaction of Facebook and Twitter (later also Amazon), in terms of giving space on the web to far-right portals, was not so much a preventive or even *ex post* reaction (in the form of restrictions on freedom of expression) to the CONTENTS proclaimed by Trump himself, but a response to the EFFECTS AND RESULTS OF THE FUNCTIONING OF THESE CONTENTS (the events on Capitol Hill, where seven people were killed).

This is important in the discussion on Internet freedom. There were and are no other options for a legally justified response, because both the US and Europe still have very outdated legal solutions: the Section 230 of the Communications Decency Act of 1996 and the European E-commerce Directive of 2000. They were passed in the years when, in the name of the principle of freedom on the Internet, the platform was considered to be a kind of 'bulletin board' where anyone could put their content and where owners and administrators should not interfere.

Of course, in the process of time tools have been developed to remove illegal content or content reported as illegal or breaking certain rules. This "*notice and takedown*" mechanism was not scrupulously applied. There were many ambiguities, mainly of a legal nature, but also as a result of open fields of interpretation as to whether the content was harmful, untrue or illegal. It was not clear who was to decide on the so-called harmfulness. After much debate - and despite resistance, including from Internet companies, pointing to various technical problems with content management and filtering and to the problem of how quickly legally (or socially?) unacceptable content should be removed - things have been sorted out in several areas. It is now clear how to deal with terrorist content (propaganda and incitement to acts), with content relating to violence against children (clear rules on definitions and how to react, although when implementing the directive there are discussions in various countries on how much time the platforms should be given to take down the content - in the Netherlands, for example, there is talk of two hours), with content relating to human trafficking, especially of women for sexual purposes (legal requirement for platforms to react) and - for all the controversy surrounding the solutions adopted - with copyright issues.

European disputes over the shape of the Copyright Directive (mainly around the initial Article 13, and now Article 17) revealed an interesting premise for thinking about the removal of illegal content (content covered by the law, but obtained and promoted in violation of the law). Publishing companies and legislators (finally) recognised that in order to limit the illegal proliferation of content, that flows through the Internet, it should be filtered before it is published online. Isn't this filtering requirement - necessary for protection purposes (for copyright) - turning into a kind of censorship? Especially in all cases of difficult interpretations (e.g. when content is processed in a grotesque, satirical way, in the misrepresentation of content in order to mock it artistically and semantically), where suddenly platforms (companies) would have a tool to stop the circulation of some content. A measure of the difficulty of the latter issue is the need for the European Commission, in cooperation with partners, to present a kind of guide on how to apply the provisions of Article 17 in practice (it is expected that this guide will appear in February 2021).

It is hypocritical to say that the large platforms must be limited in their power while, at the same time, expecting them to take on the burden of defining what is illegal and what is not compliant in some cases.

It is very wrong to expect the technology business to determine the shape of autonomous legal solutions. And precisely to avoid this, there is now a whole discussion about Internet freedom and the tools and conditions (premises) under which platforms must and should take responsibility for published content. This is particularly important at a time when the Internet is used on such a scale.

Internet access and network-telephone (mobile) activity is already benefited by 5.2 billion people, and 4.14 billion people participate in social media (Facebook - 2.7 billion, YouTube - 2 billion, WhatsApp - 2 billion, Twitter - 360 million) [3]. About 2 million people join the users of digital tools every day. It is clear that the new rules sought for the social functioning of the Internet must be global in nature. In this context, it would be good if the rules on privacy (European Data Protection Regulation, operated since May 2018) and cyber security (standards and certifications) become universal. Just as it would be important to spread worldwide the rules related to the responsibility of platforms for the aforementioned issues when it comes to the operation of certain content online (terrorism, violence against children, human trafficking, and copyright under reliable basis).

So what remains sensitive and dramatic in terms of social order when we talk about the presence of certain content in messages posted on platforms or social media? Two incredibly important issues: DISINFORMATION and the propagation of untruths (*fake news*, *fake science*, including conspiracy theories) and the HATE SPEECH, hurtful speech in its various manifestations.

When looking for appropriate legal solutions concerning the responsibility of platforms, it is worth seeing who the sources of disinformation and hate speech are. Here we enter a completely new sphere - the disinformation industry created by government administrations, the world of politics and services commissioned by them. This is the reality of IMPROVED MANIPULATION and the WEAPONISATION of PUBLIC COMMUNICATION. Thus, it is a great strengthening of giving politics a violent character, intensifying violence against the social order (e.g. exhortation under false slogans not to be vaccinated, which is a violation of the principles of public good in the name of imaginary and extremely individualised freedom), and finally - a huge support for creating conditions for extreme polarisation using all techniques of extreme, tribal divisions in society. In the era of populism, the latter becomes an efficient tool of governance.

Research conducted by the Oxford Internet Institute [4] shows how fast the number of countries where large-scale manipulation communication is used is growing - in 2020, it was already happening in 81 countries, including Poland. Different channels and different forms of social media activities are used: government agencies, politicians and political parties, private contractors (special agencies, of which there were almost 70 in the world in 2020), pseudo civil society organisations (e.g. such as Ordo Iuris), citizens themselves and network influencers. You can see how comprehensive these impacts are. Only nine countries make use of all of the indicated possibilities; Poland is among them (between: Israel, Kuwait, Libya, Malaysia, the Philippines, Russia, the UK and the US - an otherwise puzzling list). The authors of the report write about cyber centres, cyber divisions, and cyber warfare. Real (often hacked) and fake accounts are used to spread disinformation. Sometimes they work in an automated way, sometimes bots dominate, sometimes real people are involved (their task is to develop interactions with those who are being attacked, slandered or who express a different opinion in public). Trolling is used, for example in actions to discredit the opposition. This was done, for example, in Tajikistan, where the Ministry of Education hired teachers, and it was done in Poland, where employees of the Ministry of Justice were hired to manipulate and defame judges.

Part of this war takes place through advertising. The European Commission's announcements show an intention to subject advertising (including political advertising) to greater control, although it is not yet clear how this would be done. However, there must certainly be greater transparency in advertising (information about the sender, recipients, funding, if the ad is political, and confirmation that the content presented is advertising).

The goals of this digital war machine are unfortunately obvious. They focus on pro-government propaganda and manipulation, attacking the opposition and political or ideological opponents, suppressing genuine social and civil activity through intimidation, and on the extreme polarisation of society.

This, of course, creates a great danger for a democracy understood as the participation of citizens in decision-making processes under transparent rules, a democracy understood as fulfilling the rules of law and a democracy that respects difference. It is such a democracy that creates opportunities for dialogue and deliberation on important matters. When I hear that the Internet is currently the greatest threat to democracy, I think that this is a mistaken assessment. The threat is populism and the people who create it, and also - from a slightly different perspective - those who give in to populism and vote for populists. Traditional, extremely tabloid media are also a model in a negative sense. A study conducted by a research centre at Harvard University [5] on Trump's presidential campaign showed that the Internet was secondary, with the main messages of negativity generated on Fox News, for example. Unfortunately, the Internet gives efficient tools to populism (is it more efficient than radio for Hitler in the 1930s?).

And this is due to the nature of the Internet, which has created an 'attention economy' on a massive scale. All the rules of the Internet focus on the fact that our attention is constantly being generated, manipulated, valued and degraded [6]. However, this should not lead to opening legal and practical gates to "close" the Internet, restricting access to it on a mass scale. In the work on Polish cyber-security legislation, there was such a theme (although it seems to have disappeared now?) that the authorities could "close the Internet" for reasons of national security... This is dangerous for democracy and freedom, as confirmed most recently in early 2021 by the decisions of the generals who staged a putsch in Myanmar and "closed" the web (access to it) for several days wishing to stop organising protests.

Platform administrators removed more than 317,000 websites and accounts from the web in 2019 and 2020 [7], but this did not put a dam on disinformation, because it could not. Although, on the other hand, it must be acknowledged that the principle introduced in 2017 as part of the European Union's policy (under the auspices of Commissioner Vera Jourova) of cooperation between partners: governments in some countries, the European Commission, internet companies and civil organisations, based on a common Code of Good Practice for responding to disinformation and, above all, hate speech, has begun to bring results. And it is resulting in an increasing scale of taking down content reported as "false" or "hateful". In December 2020, Commissioner Jourova announced that there will be a new version of the code in spring 2021, with stronger obligations for parties to respond more quickly and effectively to lawbreaking, but also to what destroys social order and many people's personal sense of order [8].

That is why a debate about freedom on the web - which would take into account that the freedom of one must not destroy the freedom of another in the public space of the web - is crucial.

Unfortunately, we live in a world turned upside down.

And at a time when in the European Union important work is being carried out on a fundamental reform of the eCommerce Directive, that is the regulation on digital services [9], where the issue of responsibility of platforms is one of the most important - in Poland the government (Minister Ziobro) proposes its own solutions. And it is a smokescreen - because they talk about freedom, while in the meantime it is about enslavement. In various countries, the slogans "security" or "freedom" are used to actually restrict civil rights.

According to an analysis conducted by the International Press Institute [10], in recent years, and even months, in many countries, under the pretext of fighting disinformation, laws have been implemented that actually restrict freedom of expression, including journalists (this includes criticism of governments). The alleged defenders of freedom of speech and 'real truth' are political leaders such as Putin, Orban (the penalty for disinformation in the COVID-19 emergency law in Hungary can be up to five years in prison), Duterte in the Philippines, Erdoğan in Turkey, Ortega in Nicaragua, Bolsonaro in Brazil or, of course, Lukashenko in Belarus.

The Polish Panoptykon Foundation [11] has rightly called these proposals "surveillance law 2.0."

The first surveillance law can be called what came into being in 2016, as part of a skewed interpretation and implementation of a European directive, the so-called Police Directive, created as part of the work on data protection regulation. It concerned, among other things, enabling internet surveillance by introducing the so-called "fixed link" between companies (which serve users and their online presence) and services (the police and eight other services). The European solution does not talk about a fixed link. It limits the extraction of information to specific crimes (the five most serious crimes as interpreted by the Court of Justice of the European Union). For Polish secret service coordinators, any crime would be dangerous and serious. And, crucially, in the European proposal, decisions on access to Internet user data and content should be preceded by *an ex ante* evaluation by a court. Until now, admittedly, user data could be stored by communication platforms and channels at their own discretion as to how long they should be kept. But in the new draft, Minister Ziobro has included a provision that data must be retained (stored) for at least 12 months. Similarly, as it is applied in Poland with telecommunication data. Meanwhile, according to the October ruling of the European Court of Justice on retention - this is unauthorised and illegal!

The intention of the drafters is to create a (supposed) redress mechanism for users when their content is taken down from websites or social networking sites. Rules of appeal are indeed necessary, and this is what the European Union is working on, but they must be preceded by clarity of content moderation rules (justification of why something should be taken down) - something that is being discussed as part of the work on digital services regulation. Meanwhile, Ziobro, ahead of his time, says: challenging moderation decisions in Poland will be able to be directed to a new institution, the Council for Freedom of Expression. And this Council would be created without a credible civic and political consensus, meaning it would clearly become a weapon of the ruling party. If the Council challenged a moderation decision of a platform, and a user disagreed, they could go to court.... administrative one! Which in advance delays the case for years and shifts the burden of the dispute into the sphere not of legal protection of civil rights, but of the functioning of administrative procedures. And finally, the key point: the lack of confidence in the Council and in the whole procedure. Thus, it looks more like a desire to defend content that may be considered Nazi, when platforms would be ready to take it down in accordance with the letter and spirit of Polish law and considering social order as an important value. Or a willingness to defend - in the name of freedom of opinion - the removal from the platforms of content that demeans LGBT plus people.

However, in order to be able to effectively and transparently curb hate speech and disinformation - including through effective platform accountability - a few things need to be sorted out.

Firstly, to define precisely the rules of operation and the responsibility of the platforms in situations where such content appears. A new formula of the "notice, report and take down" model is to be created - with rigorous response times and a description of all the necessary procedures, as well as penalties for not fulfilling these obligations; in the Commission's proposal, the penalties would be even higher than for breaking privacy rules.

Secondly, to have clarity (in the European Union and in all member states) on the legal definition of what is illegal, breaking the law and requiring removal from the network circulation. Commissioner Jourova announced [12] that in 2021 the European Commission will take the initiative to expand the list of offences (under Article 83(1) of the Treaty on the Functioning of the European Union) and criminalise hate speech activities and those leading to hate crimes. This ordering of the law is crucial in order to take down content deemed to be illegal without an excess of interpretation problems. Harmonisation of the law in these areas across the Union is important because the Internet cannot be 'closed' within national borders. At present there are many differences, for example in the criminalisation of words and actions that demean people of a different sexual orientation or people with disabilities.

Thirdly, to find a way to limit, without violating the rights of freedom of expression, the free circulation of "harmful content" on the Internet. After discussing these issues for a whole year (2020), the European Commission's concrete legislative proposal already focuses exclusively on removing and restricting the presence of what has been defined as illegal online. Thus, doubts and areas of ambiguity remain - is some variety of false information a violation of the law, or an expression of someone's interpretation of an issue? This is the most difficult challenge. Because - who would define the harmfulness of this content, and in an impartial way? Often, nuances may determine whether or not something is considered harmful; and this may result from different sensitivities, arising from the cultural patterns of a given social, religious, or national group.

Fourthly, understand that an appropriate legal framework (such as the new digital services regulation) is only a part of solving the problem. Because the quality of the new offer in the treatment of these matters will be determined by the real enforcement of the law. Will there be a code of inevitable punishment for posting content calling for Nazism or physical destruction of a person with different views, or for causing psychological damage to a person harassed and humiliated online, preceded by actions of the police and prosecutors, initiated in some cases ex officio?

Fifthly, we need to ensure that the appeal mechanism, especially in cases that may give rise to controversy and differing interpretations, is well-described, is based under the principles of consumer rights and is efficient in day-to-day operation. It would be good, however, if the European Union were to draw up a list of good practices in matters of dispute concerning the understanding of the meaning and purpose of the operation and impact of some harmful content, such as disinformation, because it does not appear from the legislative proposal that such content should be treated as illegal and subject to mechanisms for removal from the network.

Sixthly, to promote and ensure a sense of security for platforms or their administrators, who will on their own initiative carry out actions to track down, identify sources, restrict access to or remove specific illegal content (the 'Good Samaritan' principle) or content reported as harmful or dangerous by individual users or collective entities (there is a proposal for a system of trusted *flaggers*).

Seventhly, to conduct moderation transparently and to set moderation standards, also to pay attention to differences and problems resulting from the different moderation techniques used - whether algorithm-based or human-led (all in conditions in which the draft EU regulation rightly confirms the lack of obligation for universal content monitoring on platforms).

This last point requires in-depth analysis, especially now, during the debate on the Digital Services Regulation.

It may turn out that simple rules on the removal of illegal content will not solve the problem of content bordering on freedom of expression and infringement of someone's goods and freedoms or violation of social norms. In order to minimise the threats from the Internet to the freedom of those who may be painfully affected by words and attitudes spread on the Internet, there is a need for a growing awareness of these issues and threats, as a result of ongoing public discourse and education. But the focus of attention and legislative effort in this area should not only be on content that is illegal or exceeds ethical standards (to remove it from circulation), but also on strong guarantees of freedom of expression. This is why content moderation is so important [13].

With the scale of content flow on the Internet as it is today, there is no other option but to use Algorithmic Content Moderation Systems (ACMS) for moderation. They make moderation at all possible, but at the same time they have various limitations: they are "blind" to contextual understanding, they function poorly in less popular languages, and they have low sensitivity to content creativity. Moderation schemes depend on the nature and model of the algorithmic system and the quality of the data on which it is based ("trained"). This raises an obvious and important question: what is, and what should be, the control and supervision of moderation models? The EU project assumes the participation of people in supervision; it also stresses the importance of the principle of transparency, i.e. making the mechanisms of the system publicly available. Moreover, some believe that in order to guarantee freedom of expression, the functioning of the ACMS should be subject to an impact assessment prior to the start of work - a Fundamental Rights Impact Assessment (FRIA). All this to avoid a situation in which the lack of transparent rules for analysing content (as far as possible) would, on the one hand, threaten to restrict freedom of expression and, on the other, prevent the swift removal of what is illegal or in breach of the social and democratic order.

Disputes over freedom of expression on the Internet are important for the quality of democracy and the full guarantees of the rule of law.

They cannot be trivialised by turning them solely into a dispute with large technology companies. And they cannot be trivialised by placing the freedom of expression above the rules of social life in terms of the norms of coexistence with others and respect for other people's freedom and values. Neither can they be separated from the broader context in which the power of populism in contemporary politics and the involvement of states, governments and politics in disinformation wars, manipulation and the generation of hatred and social divisions are becoming increasingly evident.

Finally, there should be no illusion that the legal framework will, by itself, solve all the problems, because beyond it, beyond defined illegality, there is a range of issues, behaviours, content. And this (removal of "destructive" content) will require codes of cooperation between companies and trusted civic organisations and new institutions at the European Union level (such as the proposed European Board for Digital Services) and in member states (DSC - Digital Services Coordinators, competent institutions established in each country, with a strong position to resolve controversial issues). Will it pass the practical test? We do not know, but we have to try.

This seems to be the only way to ensure that the dispute over freedom on the Internet wins out over freedom and the lasting protection of individual rights.

Michał Boni

February 2021.

REFERENCES:

[1] Much of the information about events around the Capitol is provided by: *Subverting the Election. How a Presidential Lie Stoked the Attack on the Capitol?* , "The New York Times", 6-7 February, 2021.

[2] The need for such regulations and, more broadly, the various dimensions of a safe Internet were recently written about for the Civic Institute by Ania Mierzyńska, *In a hybrid world we need a socially safe Internet. Five challenges of the digital revolution*, Civic Institute, electronic version.

[3] The data for: Digital 2020 October Global Statshot Report, October 2020, in cooperation with Hootsuite and We Are Social.

[4] Following the report: S. Bradshaw, H. Bailey, Ph.N. Howard, *Industrialized Disinformation. 2020 Global Inventory of Organized Social Media Manipulation*, Oxford Internet Institute 2020.

[5] *US Elections Disinformation Tabletop Exercise Package*, The Harvard Berkman Klein Center Study, 2020.

[6] An interesting article on this subject, discussed in fact for a long time: Charlie Warzel, *The Cassandra of the Internet*, "The New York Times", 6-7 February, 2021.

[7] The data for: S. Bradshaw, H. Bailey, Ph.N. Howard, *Industrialized Disinformation...*, op. cit.

[8] Vera Jourova, Commissioner, the EC, on the occasion of the publication of: On the European Democracy Action Plan, Brussels, COM(2020) 790 final.

[9] Draft Digital Services Regulation, 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, Brussels, COM(2020) 825 final.

[10] International Press Institute [w:] *Inconvenient truths*, "The Economist", February 13th, 2021.

[11] Very good review: a *Freedom of Speech Act? More like a surveillance law 2.0!!!*, Panoptykon, article dated 4 February 2021, https://panoptykon.org/inwigilacja-2-0

[12] Vera Jourova, on the occasion of the presentation of the document EC: On the European Democracy Action Plan, Brussels, COM(2020) 790 final.

[13] I use many of the suggestions after: Sunimal Mendis, Tilburg University, presentation on the determinants of freedom of expression, Digital Humanism TU Wien, Lecture series, Freedom of Expression in the Digital Public Sphere, 2 February, 2021.

Michał Boni - Doctor of Humanities, activist of the underground "S" and head of the Mazovia region of the "Solidarity", advisor in Jerzy Buzek's government, member of the Polish Parliament and minister in many governments, head of strategic advisors to Prime Minister Tusk and the Permanent Committee of the Council of Ministers, founder and head of the Institute of Public Affairs, author and coordinator of work on the report POLAND 2030 and reports YOUTH 2011 and YOUTH 2018, the first Minister of Digitalisation in Central and Eastern Europe, author of the programme DIGITAL POLAND, Member of the European Parliament; Currently assistant professor at SWPS (University of Humanities and Social Sciences) and Senior Researcher Associate at the Martens Centre (think tank in Brussels), member of the Advisory Board of the Digital Enlightement Forum, senator for SME Europe, head of the Programme Council of the FISE Foundation; author of many articles and papers.