ALLIANCE FOR HEALTHY INFOSPHERE

INPUT TO OHCHR UPCOMING REPORT ON DISINFORMATION AND HUMAN RIGHTS

The disinformation phenomenon impacts human rights in a variety of ways. On the one hand, implementation of digital space regulation is a valid need, on the other hand, in many countries, such laws are often used to silence dissenting voices. Such practices were particularly numerous in 2020, as states restricted flows of certain information in an effort to mitigate the COVID-19 infodemic or under the disguise of doing so.

Alliance for Healthy Infosphere, composed of 11 civil society initiatives, private companies, think tanks and academic institutions across 6 European states, advocates for a healthier information environment for all, while preserving democratic principles and human rights, one of the most important of them being the right to free speech. Below are our joint inputs to UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression into her upcoming report to the Human Rights Council:

1. **What do you believe are the key challenges raised by disinformation? What measures would you recommend to address them?**

Challenges:

- **User manipulation & subversion of democratic processes**: Disinformation campaigns are often deployed as a tool to influence voting behaviour during elections and they can also negatively interfere with other democratic processes. Furthermore, personal data of social media users harvested through digital platforms can be misused for micro-targeted spread of disinformation to achieve voter suppression, for example. The vulnerability of the digital platforms to data misuse has been exposed in the Cambridge Analytica scandal in which a private company harvested user data to influence voter behaviour in the US.
- **Information chaos, social polarization & radicalization**: Digital revolution brought with itself opportunities for wider populations to access more information sources as well as to produce content themselves. However, another consequence of this change is also proliferation of disinformation and hateful content disseminated via social media which contributes to information chaos, growing polarization and radicalization of societies. Typically, disinformation campaigns are very good at exploiting already existing social cleavages and further amplify them. Furthermore, the corrosive effect of disinformation and conspiracy theories can lead to total alienation from facts and reality, and can cause violence (such as genocide in Myanmar). Public opinion polls provide the impact of disinformation and influence operations on the public - for example, in 24% Central and Eastern Europeans think that COVID-19 vaccine is a tool to implant nano-chips and control people.
- **Unnecessary loss of life**: Disinformation and conspiracy theories have led to deaths, particularly as conspiracy theories and health hoaxes about the COVID-19 pandemic flooded the information space.
- **Disinformation is a profitable business**: Social media platforms and other involved actors profit from disinformation and conspiracy theories. Because disinformation is now a self-sustainable economic ecosystem, it is very difficult to disrupt.
- **Lack of transparency & ability to enforce community standards systematically**: Social media platforms are incapable of enforcing their own rules of conduct in various countries and the way their community standards are being applied lacks transparency.
- **Trust in the mainstream media and media sources is impacted:** Disinformation narratives claiming that 'mainstream media cannot be trusted' strive to undermine the trust of citizens in investigative journalism,

one of the cornerstones of functioning democracies. However, when thinking about disinformation, it is important to stress that disinformation is often promoted through traditional information channels such as TV and mainstream media in many states as well.

- **Disinformation & hate speech online negatively impact rights of minorities and women**, as their rights in many countries are being systematically impacted by disinformation campaigns demonising gender equality, LGBT+ rights, ethnic minorities, etc. In the past years, liberal policies and actors have been labelled as a threat to traditional values and national identity in many countries.
- **Geopolitical impact:** Disinformation is often used as a tool in influence operations conducted by foreign states with serious security implications.
- **Economic**: Viral hoaxes have led to stocks' fluctuation and hence, can have serious financial implications. For example, in 2013 a viral hoax about an alleged explosion in the White House caused intense impact on stock exchanges.
- **Research issues as a consequence of non-existent transparency**: Disinformation phenomenon is difficult to research as digital platforms often do not provide researchers with adequate data and information necessary to conduct meaningful studies. As researchers are not notified about planned algorithmic changes of digital platforms, it is nearly impossible to establish conditions which would guarantee veracity of findings. Without solid scientific evidence it is impossible to truly evaluate the impact disinformation has on our societies, as well as to suggest truly efficient solutions.
- **Algorithmic non-transparency:** Users increasingly consume information without choosing the source themselves, as mechanisms indexing content personalise the sources based on information which they collect about the user. This results in reduced agency in the user's decision making. This further exacerbates the problem with studying disinformation because algorithms indexing content for each user in a personalised way operate as a 'black box', meaning that only the designers of the system are aware of how it really works. In the meantime, users are left in the dark, while it is nearly impossible to study how algorithmic indexing contributes to the spread of disinformation.

Recommendations:

- **Media literacy and critical thinking**: Education systems need to be reformed to entail the digital skills and responsibility. Such education needs to start at an early age. Finland could serve as a successful example showing how resilience to disinformation can be fostered through an agile and modern education system. Life-long learning and other special programmes should be established as well for older generations or disadvantaged communities.
- **Digital platforms regulation:** Digital platforms need to operate in a clearer legal framework. As voluntary self-regulation measures applied by social platforms have not worked, efficient digital regulation/ co-regulation regimes, as currently being developed by the EU, will need to be implemented. There are other promising models which would help address the situation such as updating digital platforms to information fiduciaries.
- **Application of community standards**: Rules of conduct need to be enforced systematically, regardless of the size of a country. Social media platforms need to invest into local capacities for fact-checking and content moderation. Furthermore, functional independent appeal processes need to be set up.
- **Demonetize disinformation**: Tech giants need to do more to demonetize disinformation. One of successful examples is an appeal to advertisers to stop placing ads on disinformation sources.
- **Transparency**: We need greater transparency of algorithms and independent audits in order to see whether the measures taken by social media platforms are working and effective.
- **Invest in independent journalism and develop strategic communication** of public institutions.
- **More data**: More detailed country information on malign coordinated bot activities, trolls or fake accounts or how much malign content was taken down by social media platforms is required in order to gain full scope of the problem. Similarly, more detailed information on advertisements on social media platforms are needed to identify who is being targeted and why as well as to identify sources of funding.

▶ **What is illegal online / offline:** Often existing legislation on illegal content is not systematically enforced. Capacities to prosecute such digital offences should be further developed and funded, for example through taxation of digital platforms.

2.

a) **What legislative, administrative, policy, regulatory or other measures have governments taken to counter disinformation online and offline?**

Legislative, administrative and regulatory measures adopted by states greatly vary and those adopted by non-democratic states or backsliding democracies and hybrid regimes often cause more harm than benefit, for example:

▶ **Misusing multilateral institutions for nefarious interests:** Efforts to 'regulate' cybercrime by Russia and China at the UN level could pose threat to dissenting voices (see controversial UN resolution on cybersecurity). Attempts to address 'the use of information and communications technologies for criminal purposes' which also includes the category of 'fake news' by non-democratic states raises serious concerns about the impact on human rights such as freedom of speech and access to information.

▶ **Disinformation as an excuse to suppress dissent:** Some governments used legislation and other measures to silent the opposition, media, NGOs and strengthen their control of media or social media platforms to spread their propaganda - state capture of media via ownership, firing investigative journalists, re-division of money or ad revenues among the state- controlled media, refusal to renew broadcasting licenses.

▶ **Half-baked digital laws with questionable impact:** regulating the digital sphere is an incredibly complex endeavour and the effort to address the issue of disinformation has sometimes led to passing of controversial laws, which can impact freedom of speech.

▶ **Inefficient digital laws:** Regulation of digital space can also in practice be inefficient. German Netz DG law, for example, did not so far cause the impact which was originally envisioned for it, potentially also due to digital dark patterns deployed in digital platforms' design which prevent easy user reporting of hate speech.

Promising approach to regulation, contrasting with the examples above, is ongoing:

▶ **On the EU level:** Robust discussions and proposed legislation about the needed shape of digital space as represented by Digital Services Act (DSA) and European Democracy Action Plan (EDAP) are exemplary efforts to shape digital space in a democratic way without impacting the freedom of speech. This could lay groundwork for global standards.

▶ **In the UK:** Thorough analysis of the online harms as summarised in the Online Harms White Paper will lead to proposal of legislation with established Duty of Care for digital platforms to 'keep users safe and tackle illegal and harmful activity'.

b) **What has been the impact of such measures on i) disinformation; ii) freedom of opinion and expression; and iii) other human rights?**

▶ **Lack of transparency complicates measurement of regulations' impact:** Impact of regulation implemented in democratic states is difficult to fully grasp as social media platforms still place significant barriers of data sharing for research purposes.

▶ While adoption of the voluntary Code of Practice of Disinformation by tech giants has led to more transparency and increased efforts to fight disinformation in the European Single market, the progress has not been sufficient. According to Commissioner for Justice, Consumers and Gender Equality Věra Jourová, Commissioner for the Security Union Julian King, and Commissioner for the Digital Economy and Society

Mariya Gabriel: '*While the 2019 European Parliament elections in May were clearly not free from disinformation, the actions and the monthly reporting ahead of the elections contributed to limiting the space for interference and improving the integrity of services, to disrupting economic incentives for disinformation, and to ensuring greater transparency of political and issue-based advertising. Still, large-scale automated propaganda and disinformation persist and there is more work to be done under all areas of the Code. We cannot accept this as a new normal.*'

▶ **In authoritarian or illiberal democracies measures to 'fight disinformation' are often used to further spread false narratives, propaganda and to <u>suppress dissent</u>.**

▶ In Bangladesh for example, the Digital Security Act (2018) has been used to <u>charge over 500 people</u>, many of them being human right defenders and journalists. The law is often used against those who criticise governments' handling of the Covid-19 pandemic.

▶ Similar <u>cybercrime law</u> has been passed in Egypt where it allows state authorities to censor websites and content 'poses a threat to national security or compromises national security or the national economy'.

▶ In Russia, the new law on <u>Covid-19 disinformation</u> has been used against journalists criticising governmental response to the health crisis. Under the law, individuals face up to 23,000 EUR in fines or up to 5 years in prison, while media outlets can be fined up to 117,00 EUR for spreading what state authorities deems as 'disinformation'.

c) **What measures have been taken to address any negative impact on human rights?**

▶ In illiberal states, no measures to counteract the negative impact of digital legislation on human rights were taken, because these laws are in many cases designed in a way which deliberately limits freedom of speech.

▶ DSA and EDAP initiatives on European level attempt to find the fine line between a healthier information space and freedom of speech which is why the difference between disinformation & illegal content is important to keep. From human rights perspective, the biggest danger in terms of regulating digital space lies in conflating disinformation (deliberate dissemination of incorrect information) with illegal content (such as hate speech in many countries, content promoting terrorism or child pornography). Treating disinformation as illegal content could lead to negative impact on freedom of speech. Instead, European approach to disinformation proposes to increase costs for disinformation producers and channelling funds to support investigative journalism & quality media outlets instead of regulating speech.

▶ Netz DG Law in Germany strived to address the proliferation of hate speech on digital platforms. However, its <u>usefulness is questionable</u> due to poorly designed user reporting mechanisms on Facebook.

3.

a) **What policies, procedures or other measures have digital tech companies introduced to address the problem of disinformation?**

Tech companies introduced some helpful measures to address disinformation, but these need to be scaled up:

▶ **Independent fact-checkers**: This is an important aspect of any digital platform's environment, however - employing 1 such person per a 2,5 million market, as is the case <u>in many countries</u>, is not sufficient. Furthermore, in many countries fact-checkers are totally absent.

▶ **Alleged increase of employees and country teams:** While responsibilities of such teams for content management is absolutely essential, there is lack of transparency in terms of digital platforms providing actual numbers of these employees publicly. Furthermore, platforms' content management reaction times have been severely hampered during the pandemic, as resources have been largely redirected towards

removing Covid-19 related disinformation, causing in return the proliferation of non-covid disinformation. For example in Slovakia, <u>Facebook's policy on hate speech is not being systematically upheld</u>, as Facebook algorithms are incapable of handling the content and not enough human content moderators are reviewing the reports.

- **Self-regulation**: While the <u>Code of Practice on Disinformation</u> is an important step forward as it introduced some improvements particularly in terms of transparency of advertising, such as the Facebook Ad Library, it has proved inefficient overall. That is why the European Commission proposed a co-regulatory framework.
- **Ads libraries introduced by Facebook and Google:** Established baseline for information on social media advertising provided is vague, enabling loopholes. Evidence also shows information provided in these libraries <u>is not always correct</u> and the libraries themselves are not present in all the markets where digital platforms operate.
- **Suspension of Donald Trump's social media accounts**: This suspension demonstrates the vital need to find balance between freedom of expression and a potential to cause public harm. It also highlights a need for systematic and fair rules to be applied across the digital platforms for all users breaching community standards (as the suspension of Donal Trump's account appeared to be rather <u>arbitrary</u>) or other terms and conditions with a possibility of an appeal and a fair review process .
- Overall, the measures implemented by digital platforms have been limited and haphazard so far, although some community standards have been updated, for example the Holocaust denying content was originally allowed on Facebook and <u>now no longer is</u>.

b)   To what extent do you find these measures to be fair, transparent and effective in protecting human rights, particularly freedom of opinion and expression?

- The measures deployed by major digital platforms are currently not enforced in a transparent, fair and systematic way. This further damages societies in an environment in which digital platforms wield almost absolute power to make decisions about what content is / is not acceptable.
- Efficiency of the measures implemented so far is difficult to evaluate due to limited access to relevant data and cooperation on behalf of tech giants, even though access to data through API and CrowdTangle, for example, has enabled researchers to conduct some analyses.
- Even justified measures may be perceived as unfair in a system which lacks proper appeal mechanisms and transparency. For example, borderline cases requiring knowledge of the local context should be reviewed by human content moderators, instead of purely AI solutions. Such decisions and their reasoning should be publicly available.
- Fact-checking is not very efficient as by the time disinformation is debunked, it often has already gone <u>viral</u>. act-checked content thus does not have equivalent reach. Fact-checking needs to be sped up by investing into additional capacities and expertise. Systematically, fact-checked content should be subsequently delivered to the same audiences which were exposed to the original disinformation.

c)   What procedures exist to address grievances and provide remedies for users, monitor the action of the companies, and how effective are they?

- Activities aiming to address grievances are still very limited, for example the newly established oversight boards, which are just starting to act and build up a precedence portfolio.
- In the meantime, redress mechanisms are very limited, often to an automated messaging informing the user that re-evaluation of a piece of content has been conducted, but the decision has not been changed. This happens even in cases of <u>harassment, cyberbullying and identity theft</u>, in which victims should be entitled to not only content removal, but also fair compensation.
- Tech giants often do not even have points of contact (PoCs) which would be available to provide cooperation to local authorities and user support whenever needed.

▶ Disinformation on traditional media is sometimes easier to address, as democratic states often have systems of independent oversight and auditing implemented. In the digital sphere however, these are currently missing.

4. Please share information on measures that you believe have been especially effective to protect the right to freedom of opinion and expression while addressing disinformation on social media platforms.

▶ **Investment in independent media, support of investigative journalism and development of skills of journalists**: Good independent investigative journalism, which takes time and requires specific skill set of journalists, is a tool to unearth facts, reveal corruption scandals and influence operations of malign actors. Investigative journalists are in competition with outlets, often funded by malign actors, that spread disinformation and propaganda and which do not uphold journalistic standards. For example, many media outlets in Serbia take over content from Russian sources, because they do not have money and enough employees to work on their own articles and conduct investigations. Therefore, investments in independent media and journalism need to be increased.

▶ **Truth and facts need better PR and better visibility on social media platforms:** Social media algorithms amplify posts with the most reactions and shares. Such posts are often full of hate, disinformation, incitement to violence and fear. Such content then quickly gathers reactions and is further shared within open and closed-door groups. Algorithms of social media platforms should instead amplify posts of verified investigative media bringing good journalism and revealing the truth. A database of investigative media and their verified social media accounts could be created to provide guidance not only to social media platforms but also to readers.

▶ **More independent fact checkers and social media PoCs**: In many countries, insufficient number or complete lack of independent fact-checkers leads to inability of social media platforms to enforce their own rules of conduct. For example, in Slovakia 1 person is covering a 2,5 million market, while in Hungary such independent fact-checkers are totally absent. In addition, social media platforms should develop a network of local representatives or point of contacts who would be closely cooperating with local public institutions, journalists and non-governmental organizations working in this field. For years the attempts of such actors to report malign and suspicious behaviour to social media platforms, bots or trolls breaching code of conduct by inciting hate and harassing users have received lukewarm response - social media platforms have either been unresponsive or it took them months to take action. Therefore, local points of contacts who would closely cooperate with local public authorities and other relevant actors should be established. This would lead to better communication between the parties and faster takedown of malign content.

▶ **Maintaining distinction between illegal content & disinformation in legislation, in line with the EU model**: While disinformation has become a buzzword within the past few years, many national or international legislations and norms do not reflect the new digital reality and it is possible to observe big differences between offline and online environments. What is illegal in the offline world should be illegal in the online environment and national legislations should be updated to reflect this. Furthermore, precise definition of what constitutes disinformation and what illegal content (such as hate speech, terrorist propaganda, child pornography, etc.) is required in order to avoid implementation of legislation which could negatively impact freedom of speech by conflating disinformation with illegal content.

▶ **Fair advertising that is disinformation free**: For many actors spread of disinformation is a business. Private companies can curb down disinformation by not buying advertisements on social media platforms or websites spreading disinformation. In Slovakia and recently also in the Czech Republic PR companies are offering their clients an opportunity not to advertise their products/ companies on platforms spreading disinformation and thus educate them to be more socially responsible in this sphere. Stop hate for profit initiatives should not last only a few months, education and close cooperation with private companies can curb down the money flow that finances the spread of disinformation and eventually destroys our societies.

▶ **Better protection and legal assistance against lawsuits**: Strategic lawsuits against public participation (SLAPPs) were identified in EDAP as a particular form of harassment increasingly used against journalists and others involved in protecting the public interest. Better protection and, if necessary, legal assistance to those actors facing SLAPPs should be established.

5. Please share information on measures to address disinformation that you believe have aggravated or led to human rights violations, in particular the right to freedom of opinion and expression.

▶ **Digital laws** passed in non-democratic/illiberal countries often <u>prosecute freedom of speech under the pretence</u> of the fight against disinformation. Many governments have resorted to these practices during the pandemic.

▶ In many countries <u>internet shut-downs</u> are used as a tool to suppress the freedom of expression, to suppress the right to assemble and the right to access information. Such measures were observed, for example, in Belarus.

▶ **Media tax laws implemented by authoritarian/illiberal regimes**:

▶ <u>Additional tax</u> on media or <u>internet tax</u> can limit the diversity of reporting, press freedom as well as the access to free and independent information.

6. Please share any suggestions or recommendation you may have for the Special Rapporteur on how to protect and promote the right to freedom of opinion and expression while addressing disinformation.

▶ Stress the need for fair rules & transparency of social media platforms. These should be applied equally no matter how big the market or country is.

▶ Develop global standards for a democratic digital space. Leaders in this area, the EU & UK already made significant progress in this respect which other countries can be inspired by and can represent a basis for further cooperation. However, at the same time we should be careful of authoritarian/ illiberal countries attempting to set the rules for information space.

▶ Initiatives and findings of the UN Internet Governance Forum could be used as a baseline for setting rules for the digital sphere.

Resources

▶ <u>GLOBSEC**:** Recommendations for Facebook, national & EU authorities</u>
▶ <u>Alliance for Healthy Infosphere: Position Paper For the Facebook Oversight Board on the suspension of Donald Trump's account</u>
▶ <u>GLOBSEC Trends 2020: Central Europe, Eastern Europe and Western Balkans at the Times of Pandemic</u>
▶ <u>GLOBSEC: Ideas for Resilient Democracies</u>

Academic:

Polarization

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." Proceedings of the National Academy of Sciences 115, no. 37 (August 28, 2018): 9216–21. doi:10.1073/pnas.1804840115.

Low trust in media

Ognyanova, Katherine, David Lazer, Ronald E. Robertson, and Christo Wilson. "Misinformation in Action: Fake News Exposure Linked to Lower Trust in Media, Higher Trust in Government When Your Side Is in Power." Harvard Kennedy School Misinformation Review, June 2, 2020. doi:10.37016/mr-2020-024.

Economic impact

Karppi, Tero, and Kate Crawford. "Social Media, Financial Algorithms and the Hack Crash." Theory, Culture & Society 33, no. 1 (May 4, 2015): 73–92. doi:10.1177/0263276415583139.

Conditions influencing resilience

Humprecht, Edda, Frank Esser, and Peter Van Aelst. "Resilience to Online Disinformation: A Framework for Cross-National Comparative Research." The International Journal of Press/Politics 25, no. 3 (January 24, 2020): 493–516. doi:10.1177/1940161219900126.

Information fiduciary

BALKIN, Jack M. "Information fiduciaries and the first amendment." *UCDL Rev.*, 2015, 49: 1183.

Evaluation of NetzDG law

Heldt, Amélie. "Reading between the Lines and the Numbers: An Analysis of the First NetzDG Reports." Internet Policy Review 8, no. 2 (June 12, 2019). doi:10.14763/2019.2.1398.

Fact checking

Hameleers, Michael. "Separating Truth from Lies: Comparing the Effects of News Media Literacy Interventions and Fact-Checkers in Response to Political Misinformation in the US and Netherlands." Information, Communication & Society, May 18, 2020, 1–17. doi:10.1080/1369118x.2020.1764603.

Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. "Fact-Checking: A Meta-Analysis of What Works and for Whom." Political Communication 37, no. 3 (October 24, 2019): 350–75. doi:10.1