Task Force 4
Digital Transformation

Policy brief

SIFTING TRUTH FROM FICTION: ENHANCED PROTECTION FROM FAKE NEWS

SEPTEMBER 2021

Muhammad Khurram Khan Global Foundation for Cyber Studies and Research
Emeric Henry Sciences Po
Paul Grainger University College London (UCL)

T20 NATIONAL COORDINATOR AND CHAIR
ISPI

T20 CO-CHAIR
IAI

T20 SUMMIT CO-CHAIR
Università Bocconi MILANO

THINK20 ITALY 2021

G20 ITALIA 2021

ABSTRACT

This policy brief examines approaches to reducing the impact of “fake news” on children, exploring elementary safeguarding measures and an educational focus to distinguish authentic from manipulative material. We focus on a two-layered approach. Our first proposal encourages social media platforms to sign up to a sharing and checking standard, involving enhanced sharing protocol and measures to encourage the production and dissemination of fact checks. We then propose a schools’ curriculum, which delineates the range and type of threat presented by web-based media, helps young people to identify these threats, provides young people with the tools to evaluate misinformation and potential deceit, and alerts them to the consequences of sharing misinformation.

CHALLENGE

Over recent years young people have been increasingly dependent on the Internet as a means of accessing information as it is convenient and accessible. However, web-based information can also by-pass the established protocols defining expertise and authenticity. Moreover, fake news is increasingly credible and pervasive, as illustrated in the context of the current health crisis, with misinformation on Covid-19 vaccines and social distancing measures widely circulating.

The pandemic has catapulted young people into a heavy reliance on the Internet. It has been invaluable in maintaining a flow of information across an isolating world, and frequently has become the only source of education during lockdown. Sadly, as with all areas of human activity, it also has the potential for exploitation and manipulation.

Children are also relying heavily on the Internet for their social interactions (MK Khan et al., 2020). The purveyors of fake news often express their information in ways that are attractive for onward transmission. Material which is sensational, salacious or subversive is easily distributed through one or two clicks. Those receiving such posts often assume, uncritically, that the credibility of the information is enhanced by association with a trusted member of their peer group.

This brief focuses on the pernicious influence of “fake news”. Misinformation, lies, and rumors are as old as history, but fake news is a cyber manifestation, often inserted into genuine news sites or appearing as a “click bait” to draw in the unwary. We therefore suggest a two-pronged approach to reducing the pernicious impact of fake news on young people, which could be supported by the G20. Firstly, we suggest that tech companies and social media platforms sign up to a sharing and checking standard. Secondly, accepting that regulation alone is insufficient to wholly adjust human behavior, we propose an emphasis on educating children, and the wider population, to enhance their ability to critically evaluate the material they receive. Those of a manipulative or exploitative disposition will always discover new scams and routes to exploitation. As things stand, the Internet enhances opportunities for anonymity, and with anonymity comes the opportunity for objectionable forms of behavior. Consequently, we look at how learning to discriminate between fact and fiction might be incorporated more widely into the school curriculum.

PROPOSAL

To sift truth from fiction and provide enhanced protection from fake news, we propose that the G20 should adopt a two-layered approach as discussed in the following sections:

ENHANCED SHARING AND CHECKING PROTOCOLS

Social media platforms are key actors in the fight against the circulation of fake news and suspicious content. However, there is an inherent conflict between, on the one hand, the necessary measures to fight circulation that lead to lower user engagement and, on the other, the for-profit objective of these businesses based on circulation and advertising revenues. Because of the complexities in achieving global agreement on regulation, we propose, *pro tem*, to set up a *sharing and checking standard*, grounded on new academic results. This standard would both describe sharing protocols and encourage the production and dissemination of fact-checks. Platforms would be encouraged to sign up to this standard to preserve their public image. Recent developments (e.g. Facebook's establishment of an Oversight Board and the G7's agreement (June 2021) on a global minimum corporate tax rate for multinationals) indicate both a developing image awareness by the digital platforms, and a growing desire by governments to establish some form of control. The time is ripe to look again at introducing voluntary standards. The Facebook Oversight Board is an encouraging approach but limited to one platform. It is designed to identify accounts that need to be suspended. Our approach is more broadly targeted at changing the practice and oversight of sharing across all platforms. They can be seen, however, as complementary approaches.

SHARING: ENHANCED CONFIRMATION PROTOCOLS WHEN FORWARDING OR DISTRIBUTING MATERIAL

Recent academic results in economics, political science and psychology have shown that simple interventions, at the stage where a user is ready to share, can substantially decrease the circulation of fake news. Pennycook et al. (2020) show that even though individuals are good at determining the accuracy of news, and in particular can distinguish fake from real news, accuracy is not a key component of the decision to share. They show that using minimal interventions, such as asking to rate the accuracy of an unrelated headline, can significantly decrease sharing of fake news. Henry et. al. (2020) show that each additional click required to confirm sharing reduces the number of sharers by about 75%.

The first element of the sharing standard we propose is to have an enhanced confirmation protocol in place when users want to forward material. In its simplest form, the user, when clicking a share button, would be brought to a new page, where she would need to reconfirm the intention to share. On this page, the user would be primed towards evaluating accuracy. These enhanced confirmation settings are easily implementable but might be resisted since they impact engagement and profits of platforms.

Such an enhanced protocol achieves several objectives inspired by Pennycook et al. (2020) and Henry et. al. (2020). First, it induces users to evaluate accuracy. Second, it increases the cost of sharing and thus helps correct negative externalities. Finally, this confirmation could also be used to at least partially eliminate sharing by bots, which has been shown to be an important factor in the circulation of fake news.

CHECKING: ENCOURAGING THE PRODUCTION AND DISSEMINATION OF FACT CHECKS

The literature has also shown the importance of providing access to fact checking in reducing circulation of fake news. Henry et. al. (2020) show in particular that this effect holds regardless of whether access to fact-checking is voluntary or imposed on the user, since those who choose to access the fact check are also those more likely to be swayed by it.

Fact-checking organizations, within or outside traditional media, have been steadily developing and have established codes of principles as part of the International Fact Checking Network. This network provides a group of trusted partners with which platforms can collaborate (Facebook, for instance, has been setting up partnerships since April 2017),[1] and could be the basis of a core group of collaborators. However, these partners need platforms for several reasons: First, platforms have direct access to the content that needs to be verified. Second, such actors are in a position to widely disseminate these fact-checking articles, and finally platforms can provide the funding necessary for the huge verification task.

This role of social media platforms entails responsibilities. The first is to provide the suspicious content to be checked to the partner organizations. This entails being transparent about the algorithm they use to identify the targeted content but also setting up a procedure for users to flag content they deem worthy of verification. The second responsibility is to systematically provide links to the fact-checking content produced by the partners as soon as it is available and to refrain from influencing the content of the verifications.

Given the amount of fake news and suspicious content that flows on platforms, a quantity that has even increased with the pandemic, some source of funding needs to be found to face the tremendous fact-checking task. It is natural for platforms to contribute to this effort themselves given that they partly benefit from this negative externality since their business model is based on engagement. The last pillar of the sharing and checking standard is thus to commit a percentage of revenues transferred to fact-checking partners.[2] It is important that the relation between the platform and the fact-checkers should not be one of subcontracting, and in particular that the platform should not pay particular fact-checkers on the basis of individual articles produced. We therefore favor the formation of a common pool of resources.

In summary, the signees to the sharing and checking standard commit to:

1. Implement an enhanced sharing protocol: It requires confirmation when users want to share content with others.

2. Send suspicious news to fact-checking partners: A transparent algorithm is needed to identify suspicious news and a procedure for users to flag suspicious content.
3. Flag news that has been checked and provide a link to fact-check: Whenever a partner produces a fact check, it must be swiftly flagged on the platform and a link placed to it.
4. Transfer a percentage of their revenues to fund fact-checking partners: This financial support could be given to the fact-checking organizations for their contributions and work

ENHANCED PERSONAL AWARENESS THROUGH EDUCATION

The National Society for the Prevention of Cruelty to Children (NSPCC) in the UK has identified the following falsities as threats to children.[3]

- Fabricated or fake news stories that might cause worry
- Viral messages containing false information can easily be shared
- Challenge videos – often, the more outrageous or unbelievable a video is, the more views and exposure are generated for the creator
- Influencers' advertising products or competitions
- Meme accounts quickly spreading unverified facts
- Opinions being shared as fact
- Abusive comments and false allegations
- Scam emails or messages sent to a personal device asking you to provide personal information or containing blackmail demands.

All involve some form of falsehood. It is very important for young people to discern trustworthy news and be equipped with media literacy skills to protect them and others from the harms of malicious online material. It is, arguably, the social responsibility of the global community to generate strategies towards enhanced protection. One of the basic skills of any education, not always taught overtly, is the ability to discriminate between fact and fiction. Evaluating evidence can be an integral part of a history syllabus and an important subject-specific skill.

Some countries are now introducing broader discrimination skills such as media literacy courses and much good practice is being developed in this area. However, in most administrations, evaluating the veracity of a statement has only been an implicit part of learning. This is no longer adequate in an age of mass media. Children are subject to a deluge of information, persuasion, and manipulation for which they are presently ill-prepared. These range through relatively harmless poor or inadequate information on search tools, to deliberate distortion of political realities, to the challenging immorality of some social media platforms (e.g. advice on self-harm and suicide). Every inbox is bombarded by hoax invitations seeking entrapment or identity theft.

In the light of these new threats in a digital age, the G20 could encourage:

1. Clear identification of the developing range of threats to children's understanding, and the gaps in protective provision.
2. Promotion of an educational syllabus for each stage of education, that empowers students to evaluate, investigate and challenge fictional or manipulative education.
3. Broadening of critical thinking skills to identify fake online content, including disinformation and misinformation.
4. Building of policies and work with the online platforms and social media outlets to alert children to such malicious online deceit.

In 2010, the UK Secretary of State for Education asked Eileen Monroe to advise on Child Protection. Although closer to children's perceptions than most, she remains focused on safeguarding as the solution. Indeed, a review of the literature shows that the majority of energy around child protection goes into regulation and prevention. This does not offer a long-term solution because those with malicious or criminal intent will find a way around regulation. Furthermore, as children mature, they will need to have developed their own means of self-protection in a world that values free speech. They cannot, and should not, be sheltered forever. Snower and Twomey (2020) in their work on Humanistic Digital Governance take a more realistic approach to the wider issue of cybersecurity. They make the point that communications on the Internet lack the social norms that constrain interactions in any normal society. Back in 2014, Shelagh McManus observed in the Guardian newspaper, "If you wouldn't do it face-to-face, don't do it online". It is important that young people are helped to re-establish these norms of human interaction in their dealings on the Internet. Fake news is not a new phenomenon but the Internet, in many ways, gives it added credibility. A study of Dutch children by Dumitru (2020) found that that both children and adolescents are not preoccupied with the trustworthiness of the information they are exposed to in social media. Disturbingly, Moneva et al., (2020), in a study of children in the Philippines, concluded "that there is no significant association between logical reasoning ability and students' vulnerability towards fake news. Some high school students do not know how to identify the credibility of information on the Internet". Paula Herrero-Diz et al., (2020) found that "empirical data, from a sample of 480 adolescents, confirmed that (1) they are more likely to share content if it connects with their interests, regardless of its truthfulness, that (2) trust affects the credibility of information, and that (3) the appearance of newsworthy information ensures that, regardless of the nature of the content, this information is more likely to be shared among young people".

There is clearly an educational deficit here. Being able to critically engage with and evaluate any unreferenced material is a requirement of mature thinking and, as such, an ability that will be increasingly important as artificial intelligence and machine learning continues to generate ever-larger amounts of material. To teach children how to spot fake news on the Internet is not a narrow and specialized focus, it is part of a broader need to have a healthy distrust of the bizarre, the uncredible and the polemical. It is also important to sift peer group gossip from more credible sources. At the World Economic Forum of 2018, the President of

the European Research Council, Jean-Pierre Bourguignon, asked the scientific community to prevail in the battle against fake news and to train a new generation of critical minds.

Increasingly the term "media literacy" is being used. McDougal et al., (2018) suggest that "policy initiatives on media literacy and media education have been growing across Europe and the English-speaking world for a few decades". Recent research at EU level has provided useful evidence on the role of informal media education and formal media education to acquire media literacy competences. They found that teaching and learning practices for media literacy education can involve various classroom-based methods most of which are based on active learning. They describe effective practice as:

- Analysis and evaluation: the capacity to comprehend messages and use critical thinking and understanding to analyze their quality, veracity, credibility and point of view, while considering their potential effects or consequences.
- Creation: the capacity to create media content and confidently express oneself with an awareness of purpose, audience and composition techniques.
- Reflection: the capacity to apply social responsibility and ethical principles to one's own identity, communication and conduct, to develop an awareness of and to manage one's media life.
- Action/agency: the capacity to act and engage in citizenship through media, to become political agents in a democratic society.

These are skills that go beyond the boundaries of cyberspace. They are fundamental capacities for survival and prosperity in a world of artificial intelligence.

Sadly, media literacy (education, training and awareness) remains, in general, insufficiently developed despite the increasingly complex, hyperconnected and polarized digital media landscape. Many administrations still do not offer media literacy programs that help distinguish between facts and opinions, and support learners to identify vicious and harmful media content. There is evidence that students better detect and discern false information after media literacy forms part of the school curriculum. For example, a pilot project in Ukraine found improvements in students' ability to identify disinformation, media biases, facts, and hate speech after being taught media and information literacy lessons as part of history, art, culture, language and literature (IREX 2020). The evaluation tested children's ability to detect, spot and analyze information in broadcast, print, and social media. The study found that the participants were twice as good at detecting hate speech, 18% better at identifying fake news stories, 16% better at differentiating between facts and opinions, and 14% more knowledgeable about the media industry. The study also showed that the participants demonstrated healthy media consumption habits and improved critical information consumption skills.

Hence the summary of our proposed recommendations is:

1. That the G20 should support the creation of a **sharing and checking standard**, as described in Part 1 and encourage social media platforms to sign up to it.

2. That the G20 should support the development of a **media literacy program**, as outlined in Part 2, for each key stage in the school curriculum and subsequently for the adult population. The G20 should recommend this learning curriculum to Member States as a move towards improving critical thinking, source evaluation and awareness of emotional manipulation.

APPENDIX

Global Foundation for Cyber Studies and Research (GFCyber) is an independent, nonprofit and non-partisan think tank. It conducts research studies, contributes policy publications, provides advisory, and intellectually indulges on various aspects of classical, contemporary, and modern cybersecurity topics. The foundation aspires to bring together experts from diverse backgrounds with key interests and expertise from the intersection of cyber policy and technology. Please visit their website for more details: **http://www.gfcyber.org**

NOTES

[1] This growing alliance now includes more than 50 partners, such as the International Fact-Checking Network, PolitiFact.com, Agence France Presse, Le Monde, and Libération.

[2] Bengani (2020) estimate that in 2019, Facebook spent about $1.2 million on fact-checking, or about 0.001% of its 71 billion revenue. This could be a lower bound on the percentage of revenues the platforms would commit to transfer.

[3] UK NSPCC, sourced 15 Feb 2021

REFERENCES

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya, "Facts, alternative facts, and fact checking in times of post-truth politics", Journal of Public Economics, vol. 182, 2020, pp. 104-123.

IREX, "Evaluation of Students' Ability to Detect Disinformation After Learning Media Literacy Techniques in School", 2020 https://www.irex.org/resource/evaluation-students-ability-detect-disinformation-after-learning-media-literacy-techniques

E.A. Dumitru, "Testing Children and Adolescents' Ability to Identify Fake News: A Combined Design of Quasi-Experiment and Group Discussions", 2020 https://doi.org/10.3390/soc10030071

P. Herrero-Diz, "Teens' Motivations to Spread Fake News" on WhatsApp, Abstract, 2020

E. Henry, E. Zhuravskaya, and S. Guriev, "Checking and Sharing Alt-Facts", CEPR Discussion Paper 14378, 2020

McDougall, J., Zezulkova, M., van Driel, B., Sternadel, D., "Teaching media literacy in Europe: evidence of effective school practices in primary and secondary education", 2018, p. 6

J.C. Moneva, R.M.N. Yaun, and I. Desabille, "Fake News: Logical Reasoning Ability and Students Vulnerability", International Journal of Academic Research in Business and Social Sciences, March 2020

MK Khan, O. Bamasaq, A. Algarni, and M. Algarni, "Fostering a Safer Cyberspace for Children", T20 Saudi Arabia, 2020 https://www.g20-insights.org/policy_briefs/fostering-a-safer-cyberspace-for-children/

M.K. Khan, S. Goldberg, P. Grainger, and B. Sethi, "Heightening cybersecurity to promise safety and fairness for citizens in the Post-Covid-19 Digital World", T20 Saudi Arabia, 2020 https://www.g20-insights.org/policy_briefs/heightening-cybersecurity-to-promise-safety-and-fairness-for-citizens-in-the-post-covid-19-digital-world/

Mueller, K. and C. Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime," Journal of the European Economic Association, 2021, Forthcoming

Munro, E., The Munro review of child protection: final report, a child-centred system. CM (8062). The Stationery Office, London, 2011

Pennycook, G., J. Mcphetres, Y. Zhang, and D. Rand, "Fighting Covid-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention", Psychological Science, vol. 31, 2020, pp. 770-780

Snower D, Twomey P www.global-solutions-initiative.org/policy-advice/revisiting-digital-governance/

Tucker, J.A., A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and

B. Nyhan, “Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature”, Technical Report, Hewlett Foundation, 2018

Zhuravskaya, E., M. Petrova, and R. Enikolopov, “Political Effects of the Internet and Social Media”, Annual Review of Economics, vol. 12, 2020, pp. 415-438

ABOUT THE AUTHORS

Muhammad Khurram Khan Global Foundation for Cyber Studies and Research

Global Foundation for Cyber Studies and Research, Washington D.C., USA

Emeric Henry Sciences Po

Professor of economics in Sciences Po Paris and research fellow at CEPR. He obtained his PhD in Stanford University.

Paul Grainger University College London (UCL)

Honorary Senior Research Associate, UCL, London and Co-Chair, T20 Task Force, Digital Transformations