SIFTING TRUTH FROM FICTION: ENHANCED PROTECTION FROM FAKE NEWS

Emeric Henry

SIFTING TRUTH FROM FICTION: ENHANCED PROTECTION FROM FAKE NEWS

1. Abstract

This policy brief examines approaches to reducing the impact of 'fake news', exploring elementary safeguarding measures. Our 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.

2. Challenge

Over recent years 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 sanitary crisis, with misinformation on COVID-19 vaccines and social distancing measures widely circulating. In this policy brief, we suggest that tech companies and social media platforms sign up to a sharing and checking standard.

3. 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. We propose 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.

4. 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.

5. 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 already set up partnerships since April 2017),[1] and could be the basis of a core group of collaborators. These partners however need platforms for several reasons: First, platforms have direct access to the content that needs to be verified. Second, they are the actors 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 to the partner organizations the suspicious content to be checked. 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 that platforms themselves contribute to this effort given that they partly benefit from this negative externality since their business model is

---

[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.

based on engagement. The last pillar of the sharing and checking standard is thus a commitment of a percentage of revenues transferred to fact checking partners.2 It is important that the relation between the platform and the fact checkers not be one of subcontracting, and in particular that the platform 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: It needs to have a transparent algorithm 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 fact-check has been produced by a partner, swiftly flag it on the platform and place a link to it.
4. Transfer a percentage of their revenues to fund fact checking partners: The financial support could be given to the fact checking organizations for their contributions and work

---

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.

References

1. Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya, "Facts, alternative facts, and fact checking in times of post-truth politics," Journal of Public Economics, 2020, 182, 104–123.

2. Henry, Emeric and Zhuravskaya, Ekaterina and Guriev, Sergei "Checking and Sharing Alt-Facts" CEPR Discussion Paper 14378, 2020.

3. Pennycook, Gordon, Jonathon Mcphetres, Yunhao Zhang, and David Rand, "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention.," Psychological Science, 2020, 31, 770–780.

4. Tucker, Joshua A., Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan, "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," Technical Report, Hewlett Foundation 2018.

5. Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political Effects of the Internet and Social Media," Annual Review of Economics, 2020, 12, 415–438.