BEST PRACTICES AND GUIDELINES
FOR FAKE NEWS MITIGATION

LINDSEY BARRETT, BENJAMIN GOH, SHANNA HOLAKO, AND YONADAV SHAVIT

27 APRIL 2017

BEST PRACTICE RECOMMENDATIONS FOR FAKE NEWS MITIGATION

I.   INTRODUCTION

These Best Practices and Guidelines ("Guidelines") were developed to assist stakeholders in the modern media ecosystem to mitigate the harmful effects of fake news on the American public. Fake news has produced a variety of different harms, including: reputational damage, physical safety issues, and the erosion of civic discourse.

These Guidelines propose solutions and techniques that we believe will empower key stakeholders to correct existing structural flaws in the ecosystem that facilitate the spread of fake news; and to prevent its spread going forward. We believe it is normatively undesirable to give the power to define what 'real news' is to any one actor, particularly the government. Our solutions endeavor to use system architecture, collaboration between stakeholders, and market ingenuity to spur creative solutions that will protect fundamental First Amendment values. Our solutions are shaped by some key principles, namely: Respect for Freedom of Expression, Accountability, Transparency, and Respect for Context. The solutions we provide, supported by these principles, will explain which stakeholders our suggestions apply to; what means those stakeholders might harness; and the ultimate goal of their efforts.

Principle 1. *Respecting Freedom of Expression*

Heavy-handed definitions and mechanisms to limit fake news, even when well-intentioned, could have the perverse effect of limiting free expression online. Our definitions and recommendations aim to protect the flourishing of a robust, free, and open Internet, while isolating and limiting the kinds of content that produce harm. In drafting policies, definitions, and coordinating efforts to halt the effects of fake news, stakeholders should take extreme care to target only what is necessary to alleviate the problem, and to avoid censorship of any kind.

Principle 2. *Accountability*

Each stakeholder should be cognizant of the role they occupy in the digital news ecosystem, particularly the responsibilities they owe to the public by taking on such a role. Publishers must acknowledge the role they play in characterizing the content they create; and platform operators must acknowledge the power they have in shaping how information is spread and absorbed.

Principle 3. *Transparency*

Transparency of process will aid stakeholders in understanding how their products are used, and will better enable the public to become informed consumers of information. While acknowledging that trade secrets and concerns about system gaming makes the appropriate degree of transparency difficult to determine, we

believe that with a nuanced approach, increased transparency is crucial for preventing problems like fake news from festering.

Principle 4. *Respect for Context*

Respect for context has two primary applications in our Guidelines. First, it speaks to our acknowledgement of differentiated responsibility between stakeholders. The entities that have held outsized roles in the creation of the problem have a commensurate responsibility in solving it. Second, it speaks to the techniques we suggest: providing greater context for information – whether from a publisher or from the platform operator – will empower readers to sort through information without any organization acting in the role of censor.

II.   DEFINITIONS

FAKE NEWS: Content that (1) characterizes false information as facts; and (2) is intended to mislead the reader into thinking the content is true, rather than opinion, satire, or parody.

PUBLISHER: Any content provider, ranging from an individual blogger to an established news outlet (like *The New York Times*).

PLATFORM: Any forum created by a non-tangible medium, such as a website, mobile app, radio station, or other portal, that can provide for the dissemination and discussion of information.

PLATFORM OPERATOR: Any entity that provides an avenue for the distribution of information. This can include a television channel, a website or social media service, a subscription service, or mobile app. The Platform Operator is the organization that administers the platform.

FACT-CHECKER: Any organization, typically not-for-profit, whose mission is to investigate claims made by journalists, institutions, and public figures in an effort to promote an objectively accurate and productive civic discourse.

III. FIVE BEST PRACTICE RECOMMENDATIONS

The following five best practices provide concrete suggestions for how different actors in the modern news ecosystem – platform operators, media companies, individual journalists, and others – can work individually, or in tandem, to mitigate its effects. The first recommendation is a technological mechanism to be created by the entity, for the entity deploying the solution. The latter four recommendations are internal priority shifts that may currently exist but ought to be better executed. Due to the heterogeneity of the modern news ecosystem, it is more important now than ever before that different actors collaborate, and work to share their expertise.

AUTOMATIC NEWS VERIFICATION

Automatic News Verification refers to a range of machine-learning tools that can be used to streamline the process of verifying and authenticating information, combined with limited human oversight.

Stakeholders & Application:

Fact-Checkers

- Begin incorporating automatic news verification strategies into your workflow, including "fake news candidate detection" tools like Politifact's ClaimBuster.[1] By working with these digital fact-checking tools now, you will provide developers with important feedback on building solutions your organization can use, and you will be able to more easily integrate new tools in the future.
- Invest in tools that search the web and identify similar claims to ones that have previously been debunked (like Hoaxy[2]), so that they can also be debunked with minimal additional effort. This can amplify the impact of a single fact check, and allow human fact-checkers to focus more on checking new claims.

Platform Operators

- Create a tool to automatically identify candidate rumors circulating on your platform using rumor metadata (how quickly and widely it spreads, who spreads it, whether it is contentious among commenters, etc.). Digital platforms have both the greatest access to metadata and the most technological capability, and thus are uniquely positioned to quickly identify potential fake news. The faster a fake claim can be identified, the fewer people will be misled and the more quickly those people will be provided with the truth.
- Create a publicly-accessible portal for disseminating these "rumor candidates" so that fact-checkers can quickly be notified of the most pressing claims to check.

All Parties

- Invest in the development of automated fact-checking technologies, such as Full Fact's "FACTS" system[3] or the Factmata Initiative,[4] and in research competitions such as the Fake News Challenge.[5]

---

[1] *ClaimBuster*, CLAIMBUSTER (April 24th, 2017), http://idir-server2.uta.edu/claimbuster/.
[2] *Frequently Asked Questions*, HOAXY (April 24th, 2017), http://hoaxy.iuni.iu.edu/faq.html.
[3] Tom Hegarty, *What is automated factchecking and how will it work?*, FULL FACT (April 6, 2017), ("...Full Fact's automated factchecking project: 'FACTS: Factchecking Automation and Claim Tracking System'"), https://fullfact.org/blog/2017/feb/what-automated-factchecking-and-how-will-it-work/.

Automatic news verification technologies, when matured, promise to be some of the most important tools in our arsenal to quickly and comprehensively counter misinformation and fake news. Investing in research and early-stage development by current news-ecosystem stakeholders is likely to pay off a great deal if, and when, these technologies are brought to bear.

CITATION STANDARDS

Citation Standards emphasize the value of providing verifiable sources for readers in a given story, article, or post; particularly through providing hyperlinked information, such that readers can independently investigate and contextualize the claims made by the author.

Stakeholders & Application:

Publishers

- Citing external sources is one of the best ways of ensuring accurate news. Media outlets should adopt guidelines encouraging judicious use of citations whenever possible, especially through hyperlinks to original sources. Providing hyperlink citations for claims allows readers to easily fact-check content themselves, and adds legitimacy to a publisher's reporting by grounding it more thoroughly in verifiable facts.

Platforms

- Incorporate "source-summary" visualizations, which use digital references (such as hyperlinks, retweets, or shares) to create a digest of the sources used in a particular article. Such digests would allow readers to understand, at a glance, where the facts in an article were coming from, and whether those sources are legitimate.
- Add "degree of citation" as an input in your news ranking algorithms, and make the way the "degree of citation" is quantified transparent. If publishers know that their content will be more visible to readers when they add citations, they are much more likely to devote resources to citing more thoroughly.

DELIVERING CORRECTIONS

Delivering Corrections encourages publishers to provide corrections for readers when false information is debunked, whether by the original publisher, or by a third-party fact-checker.

Stakeholder & Application:

Platform Operators:

- Consider designing a misinformation designation system. In its most basic form, this could entail a notification sent to users when a story they have read has been updated by the publisher through an Editor's Note or correction. Such a system could also incorporate partnerships with third-party fact-checkers. In this more expanded case, posts could be flagged with a 'disputed' tag when fact-checkers

---

[4] Dhruv Ghulati, *Introducing Factmata: Artificial Intelligence for Automated Fact-Checking*, MEDIUM (March 7, 2017) https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cf1#.ifhi5jnnd.
[5] *About*, FAKE NEWS CHALLENGE, (March 7, 2017) http://www.fakenewschallenge.org/.

reach that determination. Platforms could also retroactively notify users who have previously read content that was subsequently flagged as disputed, along with the updated status of the content.

- Another possible notification could warn users on the cusp of sharing content that that content has been flagged as disputed, whether by third party fact-checkers, or through responsive metadata. This could limit accidental spreading of false content without diminishing user agency, or giving too much discretion to the platform operator in defining false content.
- Take extreme caution in exactly how posts are characterized in any disputation system (such as when a tweet or post is flagged as disputed, a warning pop-up, or any of the other above suggestions). In most cases, it will likely be preferable to characterize a post as 'disputed by fact-checkers' or 'flagged' than to use terms with less nuance. Any 'fake news' designation, in particular, is likely to be over inclusive, and invite misuse.

CHANGING VISIBILITY

Changing Visibility encourages platform operators to consider structural changes to the fora they provide in order to promote greater transparency, disincentivize the spread of misinformation, and improve digital literacy.

Stakeholders & Application:
Platform Operators:
- Resist de facto filter feeds. If you operate a platform that relies on an algorithmic feed, consider adding controls that could allow the user to opt towards a *more* generalized, rather than more personalized, experience. Consider adding controls or tools that allow users to deliberately access ideologically diverse sources of information, or a notification that nudges users when the stories users have been reading have become ideologically homogenous.
- Release a list of the variables that are used to design the feed (but not how those variables are utilized, to prevent manipulation of the system by malicious actors).
- Design, publish, and adhere to a written code of principles explaining how the feed is designed. Make these principles accessible to the public.

ORGANIZATIONAL COLLABORATION

Organizational Collaboration encourages coordinated efforts between stakeholders in the same profession (such as among media companies or among fact checkers), and stakeholders in different professions (collaboration between a fact-checking organization and a platform provider) to facilitate better understanding between stakeholders of different roles, and to encourage cooperation and coordinated action between actors whenever possible.

Applicable to All Stakeholders
Horizontal Collaboration:
Organizations within the same sector, such as two or more platform operators, publishers, or fact-checking organizations, should endeavor to pool knowledge and resources. Smaller, less-funded organizations, such as

some fact-checking organizations, may benefit from pooled resources. Larger actors, such as platform operators or established media companies, can benefit from learning how misinformation is shared across different platforms, as between Facebook and Twitter.

- Platform operators could create an app that would notify users across platforms when information has been disputed.
- Organizations should consider convening working groups to discuss shared concerns and possible approaches to addressing misinformation.
- Platform operators could collaborate to reduce the ability of fake news suppliers to receive revenue by working together to change internal advertising policies, and pledging to work with a whitelist of advertisers who have taken steps to avoid paying for false content.

Vertical Collaboration:

Collaboration between organizations within different sectors will be particularly beneficial, given the heterogeneous nature of the fake news ecosystem.

- Publishers and platform operators can collaborate to support civic discourse, promote reliable journalism, and prevent the spread of disinformation. This could include sharing metrics that have been indicative of disinformation on one party's platform, or sharing content that has been disputed or disproved.
- Fact-checking organizations can similarly collaborate with platform operators to enhance each side's ability to combat fake news.

IV. CONCLUSION

These Guidelines offer overarching principles and concrete recommendations for players looking to mitigate the harmful effects of fake news. They establish a new standard for the modern digital ecosystem to regain public trust and promote an informed republic. Our intention is to address the roles of the identified stakeholders, and to encourage them to adopt and honor the core principles we have presented, namely respecting Freedom of Expression, Accountability, Transparency, and Respect for Context. Technology has catalyzed traditional chicanery, manipulation, and demagoguery into its modern form of fake news. But it can also be harnessed to combat those problems to preserve the fundamental benefits of a free and open Internet, in a free and open democratic society. It is our sincere hope that these guidelines will help facilitate the best possible versions of both.