

**Written Representation 167**

Name: Benjamin Goh  
Consultant

Received: 19 Mar 2018

**SUBMISSION TO THE SELECT COMMITTEE FOR  
DELIBERATE ONLINE FALSEHOODS**

**POLICY OPTIONS FOR MITIGATING HARMFUL EFFECTS  
OF “FAKE NEWS”**

**Benjamin Goh  
March 14, 2018**

## Executive Summary

Fake news is a new presentation of an age-old problem: the manipulation and misuse of otherwise innocent facts for nefarious ends. The modern information ecosystem has transformed the otherwise unremarkable proposition – that sometimes, people lie – into a complex problem without a clear solution. People increasingly get their news from online platforms, and the prioritization algorithms that determine what information users see on those platforms can skew the information they receive, with fundamentally different results from former delivery methods. A personalized feed may seem to replicate the public forum, while ultimately exacerbating political polarization by only delivering the information the user wants to see-- because that is what is designed to do. This shift in delivery methods distinguishes yellow journalism from its modern cousin, fake news: the lies are the same, but the impact is more subtle, and arguably more corrosive. Fake news articles have resulted in reputational harms; physical safety harms; the erosion of civic discourse; and, arguably, even the erosion of the democratic process. Current laws are insufficient to curb those harms, or provide remedies for them; and the constitutional protection of speech limits the feasibility of a robust statute that could prohibit the creation or dissemination of fake news. Even if such a statute were legally prudent or feasible, it would be normatively undesirable for the way it would stymie expression, and turn the Singapore government into Oceania’s Ministry of Truth.<sup>1</sup> Moreover, the heterogeneity of the modern media environment further makes a one-size-fits-all technological solution impossible.

The problem is daunting, but not intractable. I suggest a number of solutions – based on principles of accountability, transparency, respect for context, and respect for freedom of expression – that I hope will help mitigate the harms caused by fake news, without unacceptably (or illegally) limiting the Internet as a robust public forum for creative expression, political discourse, and innovation.

## The Problem

### What Is Fake News

In considering the merits of free speech and a free press, Thomas Jefferson famously lamented that “[n]othing can now be believed which is seen in a newspaper. Truth itself becomes suspicious by being put into that polluted vehicle.”<sup>2</sup> Fake news is not new, and the aversion to the media being arbiters of truth is certainly not a recent phenomenon. John Stuart Mill’s conception of a “marketplace of ideas” – where truth prevails through public discourse – accounts for misinformation as an unavoidable fixture of democratic progress.<sup>3</sup> Today, fake news fits in a broader conversation about the role of discourse in democratic decision-making.

What is different now is how news is delivered, and how extensively false information can now shape national consciousness around what is fact, and what is false. Whereas broadcast technologies once served as gatekeepers providing everyone with the same set of facts, now social networks allow “‘atoms’ of propaganda

---

<sup>1</sup> See George Orill, 1984, HARVILL SECKER (June 8, 1949).

<sup>2</sup> David Uberti, *The Real History of Fake*, COLUMBIA JOURNALISM REVIEW (Dec. 15, 2016), [http://www.cjr.org/special\\_report/fake\\_news\\_history.php?link](http://www.cjr.org/special_report/fake_news_history.php?link).

<sup>3</sup> John Stuart Mill, *On Liberty*, ROBERTS & GREEN, (1869).

to be directly targeted at users who are more likely to accept and share a particular message.”<sup>4</sup> When people obtain their news from Facebook, Twitter, and other algorithmically ordered feeds, they are receiving the information they are most likely to click on, read, and share. This is a filter for the information they receive, which prioritizes posts that are likely to stimulate user engagement. These hyper-personalized news feeds reinforce our existing opinions, rather than informing them.

The evolution of how news now circulates is symptomatic of a broader conversation highlighting society’s evolving relationship with digital technologies. Understanding how to reformulate existing approaches for the digital age is crucial to dealing with the problems created by fake news. Fake news is not just a technical problem, but “evidence of a social phenomenon at play — a struggle between [how] different people envision what kind of world that they want.”<sup>5</sup>

Claire Wardle, Research Director at the Tow Center for Digital Journalism, proposed a taxonomy to critically analyze “fakeness” in the modern information ecosystem. Her schema contains seven types of mis- and disinformation:

1. *Satire or Parody*: No intention to cause harm but has potential to fool.
2. *False Connection*: When headlines, visuals or captions do not support the content.
3. *Misleading Content*: Misleading use of information to frame an issue or individual.
4. *False Context*: When genuine content is shared with false contextual information.
5. *Imposter Content*: When genuine sources are impersonated.
6. *Manipulated Content*: When genuine information or imagery is manipulated to deceive.
7. *Fabricated Content*: News content that is 100% false, designed to deceive and to harm.

I consider types 5-7 to be particularly severe; while they have existed before, their prominence during this election cycle is primarily due to the distribution networks that have been fuelled by a click-and-advertise industrial complex. Further compounding the problem is the “atomization of media consumption” and “partisan criticism from all corners” that has eroded the trust in the media industry to truthfully report on current affairs. For the purpose of these recommended Best Practices, I define ‘fake news’ as 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. I believe that this definition parses the most harmful kinds of information from more subtle forms of expression that could be mistaken for false facts: these primary objective is to craft a definition that targets harmful content while leaving expressive content intact.

## Problems Created By The Technology

### **Physical Safety Issues**

As social media becomes more ubiquitous, what happens on social media does not stay in the digital world. Online messages shape worldviews, and then translate into actions in the physical world. Unhindered, fake news can shape people’s view of imminent threats that are not there, translating to a physical, violent response to misinformation.

---

<sup>4</sup> Claire Wardle, *Fake News. It’s Complicated.*, FIRST DRAFT NEWS (Feb. 16, 2017), <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>.

<sup>5</sup> Brooke Borel, *Fact-Checking Won’t Save Us From Fake News*, FIVETHIRTYEIGHT (Jan. 4, 2017), <https://fivethirtyeight.com/features/fact-checking-wont-save-us-from-fake-news/>.

Fake news translated into a real public safety threat when 28-year-old father-of-two Edgar Ilch drove from North Carolina to a pizzeria in Washington, D.C. and fired multiple shots from an assault rifle by way of investigating an unfounded conspiracy theory he heard.<sup>6</sup> This fake news spawned from Alex Jones, a prominent conspiracy theorist and host of a popular right-wing radio show, and claimed that Comet Ping Pong, the pizzeria, was harboring young children as sex slaves as part of a child-abuse ring led by Hillary Clinton.<sup>7</sup> Ilch plead guilty to two counts – interstate transportation of a firearm and assault with a dangerous Iapon – and luckily no one was physically injured.<sup>8</sup>

Despite being proven false, this story explosively spread and fueled Twitter, Facebook, and Instagram with more zealous attacks against the pizzeria, with similar damage resulting offline.<sup>9</sup> The owner of Comet Ping Pong, Mr. Alefantis, has spent nearly \$70,000 on security measures since the shooting.<sup>10</sup> He hired two guards to stand at the entrance during business hours, installed an alarm system and a network of cameras for both inside and outside the restaurant, as ill as a panic button to alert local police in case of an emergency.<sup>11</sup> Protests and threats continue outside Comet Ping Pong, and the hoax of ‘Pizzagate’ has spread to several other pizzerias around the country.<sup>12</sup> This episode represents how fake news an old problem with new, unexpected complications. Under the a free speech marketplace of ideas rationale, the circulation of ‘Pizzagate’ is a side effect of free expression, and truth will eventually win in a battle against falsity. But here, debunking did not squash the conspiracy theories, and rather perpetuated and augmented the fake news to be even louder.<sup>13</sup> Facts no longer change people’s minds.<sup>14</sup>

Fake news has real consequences and the shooting underscores the stubborn lasting poIr of fake news and how hard it is to stamp out.<sup>15</sup> Debunking false news articles can sometimes provoke believers to hold fast to those versions of truth by seeking out more misinformation to feed those beliefs. When one online discussion forum or thread gets shut down, it only forces fake news creators to shift their focus elsewhere.<sup>16</sup>

### **Reputational and Privacy Harms**

One of the most obvious problems created by fake news is false or manipulated facts reported about real people, causing damage to individual reputations, and individual privacy. Much of the fake news stories that have become ill known centered on public figures, such as Hillary Clinton, Donald Trump, Barack Obama or the Pope. The law allows for a wider margin of possible falsity with these kinds of figures, with the understanding that they are more frequently written about, and that seeking out a position in the public eye can required a compromise to the kind of privacy that a person would otherwise reasonably expect. But not every figure maligned in a fake news story sought the public eye, such as the pizzeria owner in the example discussed

---

<sup>6</sup> The Associated Press, *Man Pleads Guilty in ‘Pizzagate’ Shooting in DC*, N.Y.T. (Mar. 24, 2017).

<sup>7</sup> See Cecilia Kang & Adam Goldman, *In Washington Pizzeria Attack, Fake News Brought Real Guns*, N.Y.T. (Dec. 5, 2016), <https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html>.

<sup>8</sup> *Supra* note 7.

<sup>9</sup> *See id.*

<sup>10</sup> *Supra* note 7.

<sup>11</sup> *Id.*

<sup>12</sup> *Id.*

<sup>13</sup> *Supra* note 8.

<sup>14</sup> *See generally*, Dave Mass, Aaron Mackey, and Kate Tummarello, *The Foilies 2017*, ELECTRONIC FRONTIER FOUNDATION, (Mar. 13, 2017) (emphasizing the importance of defending the idea of proof).

<sup>15</sup> *Supra* note 8.

<sup>16</sup> *Id.*

above.<sup>17</sup> Fake news is often discussed in broad terms but the harms it creates can be as granular as one person's livelihood, reputation, and right to be let alone.<sup>18</sup>

### **Erosion of Democracy**

The functioning of democracy requires a well-informed and participatory *demos*, which actively contributes the marketplace of ideas to move society forward. Perspectives and ideas can only resonate, however, when they are built around a common set of objective, true facts. Without these facts, voters cannot accurately make a decision on policy issues, much less engage in constructive debate around social issues. In 2009, 11% of Americans incorrectly thought that Obama was Muslim, a factor that influenced many people's decisions in the election.<sup>19</sup> Fake news is not new in this regard, but prolonged exposure to such false information further amplifies the negative impact of misinformation, which further erodes the quality of discourse, a central pillar of democracy. In one famous example from the 2016 presidential election, fantasy news website *WTOE 5 News* falsely reported that Pope Francis broke with tradition and unequivocally endorsed Donald Trump for President.<sup>20</sup> This iteration of fake news is disconcerting as misrepresents the sentiments of an influential figure, in a way that is deliberately designed to manipulate voters.

### **Erosion of the Perception and Public Trust in Journalism as a Profession**

The constitution of Singapore enshrines a free press as an indispensable check on power. As Craig Crawford wrote in "Attack the Messenger,"<sup>21</sup> a tough press always attracts the ire of politicians who want to discredit them, but journalists can only do their job if the public believes that they are working for the public interest. Fake news can provide support for the idea that the "media" is a liberal conspiracy, or equates the quality and value of the work of established, trained journalists with that of Macedonian teenagers.<sup>22</sup> Fake news delegitimizes the appropriate role of the press in holding leaders accountable, and fulfills the accusation of corrupt politicians that the press is the "enemy of the people."

This perception of delegitimacy is dangerous, but not entirely unfounded as reputable media institutions have also contributed to the perils of misinformation. When film studio 20<sup>th</sup> Century Fox collaborated with a fake news creator to run an ad campaign for the film "A Cure for Illness," they "inadvertently fueled the online spread of made-up stories about controversial topics."<sup>23</sup> They created fake news sites with believable names, populating headlines involving Lady Gaga, President Trump, vaccinations, and mental health.<sup>24</sup> When the

---

<sup>17</sup> See Cecilia Kang & Adam Goldman, *In Washington Pizzeria Attack, Fake News Brought Real Guns*, *NYTIMES* (Dec. 5, 2016), <https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html>.

<sup>18</sup> See generally, Warren & Brandeis, *The Right to Privacy*, 4 *HARV. L. REV.* 193 (1890). In their seminal law review article, Justice Louis Brandeis and Samuel Warren enumerated the basis for a modern right to privacy, or 'the right to be let alone.'

<sup>19</sup> "Poll: Obama extends national lead over McCain". *The Washington Post*. Associated Press. (July 10, 2008).

<sup>20</sup> Dan Evon, *Nope Francis: Reports that His Holiness Has Endorsed Republican Presidential Candidate Donald Trump Originated With a Fake News Website*, *SNOPE* (July 24, 2016), <http://www.snopes.com/pope-francis-donald-trump-endorsement/>.

<sup>21</sup> Craig Crawford, *Attack the messenger: How politicians turn you against the media*. ROWMAN & LITTLEFIELD (2006)

<sup>22</sup> Samantha Subramanian, *Inside the Macedonian Fake-News Complex*, *WIRED* (Feb. 15, 2017), <https://www.wired.com/2017/02/veles-macedonia-fake-news/>.

<sup>23</sup> Sapna Maheshwari, *20th Century Fox Gives Real Apology for a Fake News Campaign*, *N.Y.T.* (Feb. 16, 2017) <https://mobile.nytimes.com/2017/02/16/business/20th-century-fox-fake-news-ad-campaign.html?smid=tw-nytimes&smtyp=cur&r=0&referrer=https://t.co/z4p5NZRxny>.

<sup>24</sup> *Id.*

entertainment industry contributes to the dissemination of fake news as an advertising stunt, they contribute to the pollution that blurs what is real and not real, taking advantage of a vulnerable public.<sup>25</sup>

The movie studio admitted their wrong and accepted fault for taking the marketing too far, but they also “revised [their] internal approval process and made appropriate changes to ensure that every part of a campaign is elevated to and vetted by management in order to avoid this type of mistake in the future.”<sup>26</sup> This type of self-correction is what I hope to emulate in suggesting self-regulatory guidelines; it also reveals fake news’ legitimate threat to the credibility and function of “real” news.

### **Evolution of the Media Landscape**

The central premise of a free press relies on two pillars: first, on the idea that society values freedom of speech; and second, that society is willing to pay for its journalistic work in a way commensurate with the public good it brings. The free press is the Fourth Estate, an important check on government power, guided by journalistic principles and ethics. In older times of print press, these assumptions do hold — civic discussions often revolved around the news and buzz in town, and everyone had a copy of the local papers to find out what is going on in the country and economy. Today, given the 24-hour news cycle, blogosphere, and atomization of press consumption, many assumptions that created a working system before start to collapse.

First, and most importantly, is the breakdown of the idea that people value a free press to the extent that they are willing to pay for it. Whereas before the press was a main source of information for people to go to about their community, today there are so many more outlets to seek the information I *want* to get. Moreover, many digital news sites make content available for free, disrupting the business model of established outlets. In this sense, if people no longer pay for content because of the expectation that they can get their news for free, the press has to reshape its practices and business models to maintain relevance and viability. For example, Washington Post announced that they will be working with Snapchat to deliver content that will appeal to people, producing “fast, visually captivating and experiential storytelling.”<sup>27</sup> Such mass appeal can rub journalistic purists the wrong way, but in today’s context of media consumption such a move is inevitable, and even prudent.

Secondly, the evolution of the media has made us confront the reality that facts do not change our minds. As research by Stanford University has shown,<sup>28</sup> “Once formed...impressions are remarkably perseverant,” and people begin to obtain a certain lens in looking at reality, choosing which facts to take in and which ones to cast out. Especially when our brains are already exhausted by the sheer amount of (social) media exposure, our brains begin taking a “short-cut to credibility” — I believe something is true when I see multiple messages about the same topic.<sup>29</sup> Facebook has become the civic town hall it was never designed to be, making its failure to stop the propagation of fake news unsurprising.

The new media landscape is worrying, because as Niels Bohr warned, “The opposite of a correct statement is a false statement. But the opposite of a profound truth may well be another profound truth.”<sup>30</sup> The discussion of fake news is inseparable from the context and medium within which the media operates. In earlier

---

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

<sup>27</sup> WashPost PR Blog, *The Washington Post will be the breaking news source on Snapchat’s Discover*, The Washington Post (Feb. 13, 2017).

<sup>28</sup> Elizabeth Kolbert, *Why Facts Don’t Change Our Minds*, The New Yorker (Feb. 27, 2017),

<http://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds>.

<sup>29</sup> Claire Wardle, *Fake News. It’s Complicated.*, FIRST DRAFT NEWS (Feb. 16, 2017), <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>.

<sup>30</sup> Niels Bohr, *Atomic Physics and the Description of Nature* (1934).

times, the community revolved around the “papers” that became the talk of the town, which created a sense of belief and standards in the medium itself (both generation of content and distribution) — everyone had a common profound truth in the poIr of civic participation through the news. Along with the participation of non-professionals into the media space, too much noise in the industry only serves to further diffuse belief in the importance of the media, providing unhealthy skepticism for fake news to thrive. Technologies may have improved to become more efficient, but humans are only just adapting to this new model of information communication. Careful correction to avoid the perils of misinformation is warranted in light of this technological shift.

## **Solutions**

As discussed above, fake news creates a wide range of societal harms that can, and should, be mitigated. The erosion of public discourse, reputational harms, privacy invasions, and physical safety issues have all been caused by fake news, and should be corrected. As has also been discussed, the proection of speech significantly limits what any statute could include. Even if I Ire able to articulate a Constitutionally defensible legislative solution to the problem of fake news, which is fairly unlikely, there are normative concerns as Ill. Beyond what is legally passable in the abstract, I do not believe that a centralized, governmental solution in the form of an enforceable statute can effectively mitigate the harms created by fake news without chilling speech to an unacceptable degree, and rendering the government an arbiter of truth.

Instead, I have identified principles I believe are crucial to upholding free speech objectives; and techniques can both support those objectives, mitigate the harms that fake news creates, and prevent similar problems from taking root in the future. The fake news ecosystem includes publishers (individual writers, as Ill as larger publication entities, such as newspapers, Ibsites and media conglomerates); platform operators (Twitter, Facebook, Snapchat, and whatever platforms may spring up in the future); and fact-checkers (such as Politifact or Poynter). I believe that articulating and spreading the objectives that publishers and content disseminators should strive for in how they approach publication will be an important part of fostering an effective solution to fake news.

## **Principles & Methods**

Misinformation is fed by obfuscation. I encourage both platform operators and publishers to strive for clarity, transparency, and accountability in their practices, so that readers are better able to discern what information they are encountering, and how they came to encounter it. I find that the most appropriate method for achieving those objectives is to provide individuals with greater context for what they read, watch, or hear. This context includes both the sources of the article’s claims, and the factual accuracy of the claims as determined by third parties. It is not enough that actors should strive to make their users informed; they should strive for those informing practices to be clearly and succinctly articulated for users, such that transparency of procedure is meaningful. The emphasis on process and accountability is a deliberate rejection of censorship, as no one entity in the media ecosystem should be the arbiter of truth. The aim is not to regulate what a consumer sees, but how they see it: with sufficient context to draw their own informed conclusions about the content. Moreover, rather than advocate platitudes in the abstract, I want to provide specific ways that different actors in the news ecosystem can incorporate these larger values into their day-to-day decisionmaking. Not every tool

will be appropriate for every actor in the ecosystem, nor combat every part of the problem, but I believe using a range of these tools will assist in combating the harms of fake news without stymying creative expression or political discourse. Below I detail a set of substantive recommendations that I believe will help mitigate the problem of false news.

### Automatic News Verification

The first step in mitigating the effects of fake news is determining which news claims are false. The traditional fact-checking process is labor-intensive, requiring checkers to find claims that may be false, parse these into checkable facts, compare these facts to existing trusted records, and if they are found to be false, to pen and publish persuasive, concise rebuttals. The swiftness with which verifiers do this is essential: the longer a claim is allowed to circulate unchecked, the greater volume of individuals will absorb falsehoods as truths. Worse, if the current news cycle passes before a claim is fact-checked, its rebuttal may go largely unnoticed by consumers. Yet existing fact-checking organizations are fairly small (PolitiFact had 10 employees<sup>31</sup> at time-of-writing, while Snopes had 16<sup>32</sup>, and factcheck.org had 8<sup>33</sup>). There are hundreds of news sites writing dozens of stories a day, numerous 24-hour televised news channels, and many thousands of users making claims on blogs and social-media. In order to properly verify all the news readers receive, I must find much more efficient ways of fact-checking. I believe that automatic news verification is a promising and effective solution to this problem.

Automatic news verification is a name for a suite of technologies that automatically aggregate, parse, and fact-check digital information. There are two approaches for a machine to verify the accuracy of a claim: (1) determining whether a claim is true in isolation, and (2) determining whether the claim is the same as one that has been previously checked. If either system finds sufficient evidence that the news is not credible, it presents the claim and evidence to a human fact-checker for final verification. This last step is crucial. Natural language processing and machine learning are a long way away from being able to reliably determine whether an article is “fake news” on their own; such a feat requires an understanding of writer tone, social context, and the implicit meanings of phrases, to the point where it has been compared to solving machine intelligence.<sup>34</sup> But by assisting humans in identifying candidate news pieces (e.g. going through thousands of articles to find mentions of “Obama birth certificate”), automatic news verification can dramatically accelerate the workflow of fact-checking organizations.

There are two primary ways to determine whether a piece of news is fake in isolation. The first is content-based. The computer basically executes all the steps a human fact-checker normally would. It parses the material into claims, compares these claims with databases of facts to determine if they are true, and aggregates the list of factual inconsistencies as “evidence.” Unfortunately, the ability to convert complex grammar into simple, checkable assertions is beyond the scope of current natural language processing techniques. However, some preliminary efforts are already underway: Factmata has built a tool to check simple statistical assertions such as “Our aid work in Somalia is paying dividends—only 0.2% of the population is severely

---

<sup>31</sup> The PolitiFact Staff, POLITIFACT, (March 8, 2017) <http://www.politifact.com/truth-o-meter/staff/>.

<sup>32</sup> Snopes.com Staff, SNOPEs, (March 8, 2017) <http://www.snopes.com/snopes-staff/>.

<sup>33</sup> Our Staff, FACTCHECK, (March 8, 2017) <http://www.factcheck.org/our-staff/>.

<sup>34</sup> Cade Metz, *The BittersJet Stepstakes to Build an AI that Destroys Fake News*, WIRED, (March 7, 2017) (“A machine that can reliably identify fake news is a machine that has completely solved AI.”), <https://www.wired.com/2016/12/bittersJet-stepstakes-build-ai-destroys-fake-news/>.



malnourished.”<sup>35</sup> As natural language processing progresses, such approaches will automate more and more of the fact-checker’s responsibilities.

The other surprisingly effective approach to determining whether a piece of news is “fake” uses the article’s metadata, completely ignoring the article’s content. The best example of this is the “propagation graph”: a machine can follow the trail of retweets, shares, and user mentions to build a network that shows exactly how a rumor spreads through social media. Work by Soroush Vosoughi<sup>36</sup> has shown that by analyzing the rate at which a claim is shared on Twitter over time, one can determine with up to 75% accuracy whether that claim is true or false. In particular, if a claim spreads more quickly early on, and then slows down as it reaches a wider audience, it is more likely to be fraudulent. Similarly, when a claim propagates from less-popular users to more-popular users, it is a good predictor that the claim is likely to be true.<sup>37</sup> Other predictive information includes the geographical location of the social media users, whether there are conflicting viewpoints in comment sections, and the reputation of the sources.<sup>38</sup> While such information may be complicated for humans to parse quickly, these sorts of analyses are precisely what machines are good at. By constantly monitoring social media propagation of claims, computers can gain reliable insight on whether to refer rumors to human fact-checkers.

This still leaves the other side of the fact-checking problem: can I detect whether a claim is similar to a claim previously debunked? Here, natural language processing generally does better (even a simple Google search accelerates the fact-checker’s process dramatically). Full Fact, a UK fact-checking organization, has been working to build a dedicated search engine that returns a list of news articles related to provided queries.<sup>39</sup> Technologically, this problem is much simpler than determining the truth of a piece of content, as pieces making similar claims will generally share word patterns. Of course, such similarity is much easier to estimate than to confirm, and the results will still be fed to a human fact-checker.

Automated fact-checking shows a great promise in making fact-checking faster, less labor-intensive, and broader in scope. Though some of the more ambitious objectives (including checking new facts) require improvements in natural language processing, many of these tools can be made possible today. I recommend that fact checkers invest more heavily in automating aspects of their workflow to leverage these emerging technologies. I also recommend that social networks provide fact-checkers with news metadata (such as propagation graphs) to quickly alert them to the stories that need to be checked. Finally, I recommend that publishers, platforms, and governments invest in automated fact-checking research like the Fake News Challenge<sup>40</sup> to continue to push this technology forward.

### Delivering Corrections

Fact-checking is only meaningful when consumers are provided with the resulting corrected information. Studies show that consumers are unlikely to go out of their way to search for the context and

---

<sup>35</sup> 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#.ifhi5jnnnd>.

<sup>36</sup> *MisinfoCon: a Summit and Creative Studio on Misinformation*, NIEMAN REPORTS, (March 8, 2017). <https://vimeo.com/205615758>.

<sup>37</sup> *Id.*

<sup>38</sup> *Id.*

<sup>39</sup> Mevan Babakar & Will Moy, *The State of Automated Fact Checking*, FULL FACT, (March 7, 2017), [https://fullfact.org/media/uploads/full\\_fact-the\\_state\\_of\\_automated\\_factchecking\\_aug\\_2016.pdf](https://fullfact.org/media/uploads/full_fact-the_state_of_automated_factchecking_aug_2016.pdf).

<sup>40</sup> *About*, FAKE NEWS CHALLENGE, (March 7, 2017) <http://www.fakenewschallenge.org/>.

veracity of their news.<sup>41</sup> It is vital that the modern news media ecosystem quickly connect consumers with the necessary context, clearly and prominently displayed, for the news they read. In the past, the traditional media has been responsible for issuing corrections when it makes mistakes. HoIver, as social networks increasingly replace traditional outlets as “distributors” of the news, the responsibility shifts to these platforms to pass on those corrections and fact-checks to readers. The applicable tools will depend on the platform in question, and differ among Facebook, Twitter, and Reddit.

Media platforms such as Facebook, Twitter, and Google know a great deal about individual users’ news-reading habits. Already, Facebook is using this information to combat fake news by alerting users that articles they have clicked on are “disputed by third-parties” if more than one fact-checking service has contested it,<sup>42</sup> and similar third-party tools exist for Twitter.<sup>43</sup> HoIver, this approach only addresses users who have yet to read/watch the news in question, ignoring the many consumers who are deceived before the facts had been checked. I recommend that when a news story is determined to be “disputed,” users who are known to have previously clicked on that story should be notified of the correction. Such a notification should be prominent and clear to ensure that users are informed about the veracity of what they had read. The more reliable such a service is, the more trust users will have in the news they see on these platforms (as they can trust that if it was fake, they will be notified).

News platforms are not the only mechanism for providing fact-checks to consumers. Certain news outlets have begun experimenting with real-time fact-check annotations for formats including transcribed political speech broadcasts,<sup>44</sup> and such annotations could be extended to any video and audio format. Of course, live-annotation requires a rapid ability to fact check. While major news networks can devote enough resources to accomplish this (e.g. in presidential debates), I anticipate that this approach will become more widespread as automated fact-checking (which is near-instantaneous) improves.

### Source Tracing & Citation Standards

The origins of an article’s claims are arguably the most important context to a consumer’s ability to evaluate an article on their own. Evidence suggests that few readers will go out of their way to check a claim’s sources or credibility unprompted.<sup>45</sup> I propose providing users with an easily-interpretable “source summary:” a trace of the claims of a news piece to their original sources. For example, a trace could show that an article by the Federalist cited a story in “The Blaze,” which cited a retlet by a public official of an account of a private citizen. Such traces would also allow users to apply their personal knowledge of the legitimacy of different sources, rather than creating a centralized arbiter of legitimacy. I believe that source tracing will foster accountability on platforms, help restore the legitimacy of the perception of bona fide journalism, and allow readers to be better aware of whether what they are reading is true.

The primary challenge in source-tracing is that many pieces of news are not Ill-sourced. Such methods are much more effective on social networks, where retlets, shares, and tags provide concrete relationships

---

<sup>41</sup> *Evaluating Information: The Cornerstone of Civic Online Reasoning*, STANFORD HISTORY EDUCATION GROUP, (March 7, 2017) <https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%202011.21.16.pdf>.

<sup>42</sup> Facebook post, Mark Zuckerberg, (March 7, 2017) <https://www.facebook.com/zuck/posts/10103269806149061>.

<sup>43</sup> *Now you can fact-check Trump’s tlets — in the tlets themselves*, WASHINGTON POST, (March 7, 2017), <https://www.washingtonpost.com/news/the-fix/wp/2016/12/16/now-you-can-fact-check-trumps-tlets-in-the-tlets-themselves/>.

<sup>44</sup> *Building an Annotation Tool on a Dime*, OPEN NEWS, (March 7, 2017), <https://source.opennews.org/articles/building-annotation-tool-dime/>.

<sup>45</sup> *Evaluating Information: The Cornerstone of Civic Online Reasoning*, STANFORD HISTORY EDUCATION GROUP, (March 7, 2017), <https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%202011.21.16.pdf>.

betlen articles and their source materials, and less effective in news articles, which inconsistently use links and other forms of citation, or forego citation altogether. The better-cited an article, the more easily its sources can be generated (and the less the consumer must rely on the accuracy of the journalist's knowledge). Of course, at some point journalists will provide facts for which they are the primary source, and I do not propose eliminating this core function of journalism. HoIver, the feIr uncorroboratable claims an article makes, and the more it can be grounded in existing fact, the more likely it is to be true. I recommend that news outlets cite source material, much of which is available online, more methodically to clearly ground their statements and place themselves above reproach.

### Changing Visibility

One standard feature of digital news curation engines is “automatic ranking:” a hierarchy, based on article content and a user's prior preferences, determining which posts and news articles are displayed most prominently to the user. These algorithms presently incorporate a wide array of features, and “estimated credibility of news” can easily be incorporated (as clickbait-iness already has been).<sup>46</sup> This approach is particularly useful because it does not require a binary “is/isn't fake” classification: the more likely the news is to be fraudulent (aggregated using the factors above), the loIr it can be ranked. The drawback to such a system is that it still puts the poIr of shaping how news circulates in the hands of the platform operators, who create the algorithms; while I encourage an emphasis on ranking credibility, there is still a transparency concern, as these algorithms are typically trade secrets, hidden from public scrutiny.

### Organizational Collaboration

The modern information ecosystem is heterogenous and complex; the way that technology has changed the dissemination and consumption of news has made it an inherently collaborative and multidisciplinary enterprise. News is written by journalists, disseminated through platforms, and funded by advertisers. I encourage collaboration betlen actors of the same sector, such as betlen publishers, and betlen members of different sectors, such as betlen publishers and platform operators. The solutions to the problem of fake news involve numerous stakeholders, none of which can overcome it alone. News publishers, news platforms, and fact-checking organizations must all collaborate to improve the media ecosystem. Below I will detail a few profitable avenues of partnership.

Fact-checkers and news platforms may have some of the most fruitful relationships. Their incentives are directly aligned: the more accurate fact-checkers can make a news platform's content, the more engaged users will be with that news, and by extension, with the platform's service. As previously described, fact checkers could benefit from early-warning signals detected by platforms over the propagation of potentially questionable news. Similarly, the two parties can work together to create a seamless “fake news notification” experience that allows users to be quickly and effectively notified of a false claim to which they have been exposed. Most people are still unfamiliar with the practice of fact-checking,<sup>47</sup> and such a partnership could significantly enhance these organizations' reach. Google's Digital News Initiative,<sup>48</sup> or Facebook's Journalism Project<sup>49</sup> and

---

<sup>46</sup> *Facebook's new anti-clickbait algorithm buries bogus headlines*, TECH CRUNCH, (March 7, 2017), <https://techcrunch.com/2017/08/04/facebook-clickbait/>.

<sup>47</sup> Eryn M. Carlson, *The Future of Political Fact-Checking*, NIEMAN REPORTS, (March 7, 2017), <http://niemanreports.org/articles/the-future-of-political-fact-checking/> (“An American Press Institute (API) study, which measured public views on fact-checking during the 2014 campaign season, found that nearly half of those surveyed are somewhat or very unfamiliar with the practice.”).

<sup>48</sup> *About*, DIGITAL NEWS INITIATIVE, (March 6, 2017) <https://digitalnewsinitiative.com/about/>.

collaboration with the News Literacy Project,<sup>50</sup> would be good examples of cross-sector collaboration. Google's Crosscheck initiative<sup>51</sup> is a good example of collaboration amongst fact-checkers.

Finally, platform operators and online content creators (as opposed to traditional media entities with an online presence) should be encouraged to join a fact-checking code of conduct, such as the Poynter International Fact Checking Network (IFCN) Code of Principles.<sup>52</sup> Facebook, for example, is requiring that the fact-checkers it employs for its platform be signatories to the IFCN,<sup>53</sup> which ensures that the standards it enforces on its platform are aligned with the standards that are endorsed by the professional news media. The old guard of professional journalism not only has the most experience in the field, but also has the best insight into how a given set of standards could incentivize future journalists and shape the direction of the Fourth Estate.

## Conclusion

The spread of misinformation online is a new presentation of When people take the information they see online seriously, very real harms can result in the physical world. Our interventions strive to uphold principles of accountability, transparency, respect for context, and respect for freedom of expression. I have taken both a technical and incentive-based eye to the solutions I propose, operating creatively within the political and legal constraints. Ultimately, I believe that these measures are an important beginning to solving the problem to deal of misinformation in the digital age.

---

<sup>49</sup> Fidji Simo, *Introducing: The Facebook Journalism Project*, FACEBOOK MEDIA, (March 6, 2017), <https://media.fb.com/2017/01/11/facebook-journalism-project/>.

<sup>50</sup> Alan C. Miller, *NLP to Help Kick Off National News Literacy Initiative*, NATIONAL NEWS LITERACY PROJECT, (March 6, 2017), <http://www.thenewsliteracyproject.org/news/nlp-news/facebook-national-initiative-accelerates-news-literacy-education>.

<sup>51</sup> Natasha Lomas, *Google and Facebook partner for anti-fake news drive during French election*, TECH CRUNCH, (Feb. 6, 2017), <https://techcrunch.com/2017/02/06/google-and-facebook-partner-for-anti-fake-news-drive-during-french-election/>.

<sup>52</sup> *International Fact-Checking Network fact-checkers' code of principles*, POYNTER, (March 8, 2017), <http://www.poynter.org/fact-checkers-code-of-principles/>.

<sup>53</sup> *Facebook has a plan to fight fake news. That's where I come in*, <http://www.poynter.org/2016/facebook-has-a-plan-to-fight-fake-news-heres-where-i-come-in/442649/> (Dec.15, 2016).