New Frontiers for Fake News Research on Social Media in 2021 and Beyond*
Invited Talk - Extended Abstract

Nir Grinberg[0000−0002−1277−894X]

Ben-Gurion University, Beer-Sheva 8410501, Israel
nirgrn@bgu.ac.il

Abstract. In the past five years, the research community made impressive strides in quantifying the dissemination and reach of fake news, understanding the cognitive mechanisms underlying belief in falsehoods and failure to correct it, and developing new methods for limiting its spread. Yet, there are many open challenges that must be addressed to ensure the integrity of the democratic process and the health of our online information ecosystem. In this talk, I focus on areas of research that are critical for advancing our understanding of fake news on social media: going beyond representative samples and convenience samples, detecting emerging fake news sources and developing new kinds of benchmark datasets for its detection, studying cross-platform impacts of platform interventions and delivering more ecologically valid experiments on social media. Progress on these fronts is necessary in order to study the pockets of society that are most heavily hit by fake news, to limit its impact, and to devise mitigations.

Keywords: Fake news · Social media · Research agenda.

In the last five years, more than 5,000 research articles were published about fake news, which stands in stark contrast to the mere 54 research articles published in the preceding five years about the topic[1]. The research community, from nearly every discipline and field, has shifted its attention to some degree to study the fake news phenomenon. Collectively, we gained new knowledge and insights on many fronts, far beyond what a single paper or presentation could cover. Thus, the goal of this talk is to focus on key challenges and opportunities that lay ahead in areas related to individuals' consumption and sharing of fake news on social media, in the development of robust computation methods for detecting fake news, and in methodology for evaluating the effectiveness of mitigation strategies. The pursuit of these promising directions would not have been possible without the foundation of new knowledge established in recent years.

Presented at the MISINFO 2021 workshop held in conjunction with the 30th ACM The Web Conference, 2021, in Ljubljana, Slovenia.

[1] Including "fake news" in either their title or abstract according to dimensions.ai [3].

1 Key takeaways from five years of research

Strictly focusing on the experience of voters on social media reveals four themes in the academic literature. First, it is evident that a lot of fake news, at least in the 2016 U.S. presidential election circulated on social media. Groundbreaking reporting from Buzzfeed's Craig Silverman revealed that the top 20 fake news stories outperformed the top 20 real news stories in terms of the number of engagements on Facebook in the months leading up to the elections [17]. Consist with that, Guess et al. show that a considerable amount of visits to fake news sites by voters originate from Facebook [7], and our work estimated that about 5% of political content flowing to voters on an average day before the election on Twitter came from sources of fake news [5].

Somewhat in contrast to the first theme, a second theme in the literature finds that the average voter saw and shared very little fake news content in 2016. Allcott and Gentzkow conclude that "the average US adult might have seen perhaps one or several news stories in the months before the election" [1]. Guess et al. report that over 90% of voters in their sample linked to Facebook data did not share any links to fake news sources during the survey period [6]. Our work estimated that the average U.S. voter on Twitter had in their Timeline only slightly more than 1% of political content coming from fake news sources [5].

A third theme "settles" the conflict of the two themes: The ample amounts of fake news on social media are concentrated in a small part of the population. Our work finds that exposure to and sharing of content from fake news sources is extremely concentrated – only 1% of voters on Twitter accounted for 80% and a mere 0.1% of voters shared 80% of it [5]. Others have noted this concentration on Facebook [6] and in online browsing to fake news sources [7], though at slightly less extreme levels. Moreover, consumption and sharing vary considerably based on individuals' political orientation, interest in and engagement with politics, and age [1, 5, 6, 7].

Finally, we now know considerably more about the psychological mechanisms behind belief in misinformation and the effectiveness of various interventions. In a recent review, Pennycook and Rand synthesize this burgeoning literature and highlight, for example, how contrary to common belief a lack of critical thinking is more dominant in falling for fake news than motivated reasoning [12]. There is also a growing body of work that describes how interventions such as attaching warning labels to content affect people's perception of veracity and intentions to share it further on social media [4, 11, 19]. A comprehensive review of this nascent area of research is outside the scope of this talk. However, it is important to note that much of this literature is based on surveys and lab experiments that decontextualize and decouple information from the social context they are naturally experienced on social media, it relies on a sample of the population that may or may not engage with fake news under normal circumstances, and that depends, to some degree, on self-reported measures rather than actual behavior.

2 New frontiers for fake news research

I focus on three research areas that can significantly move the entire field forward. In particular, I discuss challenges and opportunities in sampling, fake news detection, and experimentation.

Sampling: Given the knowledge we now have about the concentration of fake news in the population and the efforts to manipulate public opinion on social media, we need new sampling methodologies to address these issues. The rarity of engagement with fake news, of less than one in a thousand participants, makes representative surveys underpowered when it comes to studying this phenomenon, even when large surveys of thousands of people are involved. Moreover, combined with the fact that fake news sources tend to share their audience with other fake news sources [5], the inability to generate a meaningful-size sample of people consuming fake news leads to a potential blind spot – missing the communities that engage most heavily with fake news. Of course, sampling social media users based on observed behavior online can alleviate these concerns, but it brings along issues of representativeness, validity, and reliability. In particularly, this is problematic because we know that a considerable amount of political activity on social media is generated by bots, trolls, foreign actors, and other entities [9, 15], who are not eligible to vote.

To overcome these challenges, the research community needs to adopt new sampling techniques. Salganik offers two research designs that are particularly appropriate for addressing these issues [13]. The first is Amplified Asking, which refers to the use of big data to extrapolate and predict survey responses for individuals who did not take the survey. This approach is particularly useful for extending existing survey responses to the rest of the online platform. The second approach proposed by Salganik is called Enriched Asking and refers to the combination of large survey or administrative data with big observational data through the use of record linkage. By linking the responses of a set of individuals with their online behavior one can attain large samples while keeping the contamination from bots, trolls, and other accounts relatively low. Building such resources is a costly effort, but once built it can be re-used multiple times and result in many new insights about the online behavior of diverse populations.

Detection: Considerable amount of research has focused on the development of machine learning models and algorithms for the automatic detection of fake news (see Shu et al. for a comprehensive review [16]). While fully automated detection may be the ultimate goal, it is still far from substituting human fact-checkers. Most existing datasets for training machine learning models focus on veracity directly, have varying levels of granularity and definitions of fake news, and are limited in size [2, 10]. Moreover, there is little to guarantee that a model trained on the past falsehoods will reliably detect the lies of tomorrow. These are fundamental issues that machine learning models might solve one day, but they require considerably more training data and domain knowledge than current models possess.

To make steady progress, we need to work more closely with fact-checkers (rather than attempting to substitute them) and build computational models that support smaller tasks in the process of the fact-checking. For example, the research community has largely ignored the important, time-consuming, and non-trivial task of identifying claims that have already been fact-checked [14]. Benchmark datasets for this purpose are still very much absent. Another challenge that has been largely overlooked involves the retrieval of relevant evidence for supporting a given claim [18]. In addition, more research is needed to help fact-checkers direct their efforts toward claims that are feasible to check (and not popular) as well as identifying stable and persistent characteristics of credibility (e.g. audience composition of a source) that are difficult to manipulate.

Interventions: Ultimately, to reduce the spread of fake news on social media action must be taken, but the academic community cannot leave experimentation with these actions solely at the hands of platform providers. The academic community must be able to assess the effectiveness of interventions independently and free of any commercial interests. Of course, academics can try to replicate the organic experience of social media users in lab settings, but this comes at a cost to the ecological validity of the experiment and particularly its social elements. One approach that has not been sufficiently explored is the instrumentation and manipulation of social media apps and web interfaces for consenting individuals. For example, nothing stops academics from building a Twitter clone app that interacts with organic content on the actual Twitter platform while allowing researchers to introduce interventions to the user experience. Consenting individuals could be asked to use such an app instead of the regular app, and perhaps platform providers would offer such a service to the research community.

Another avenue for impactful research is the study of cross-platform effects. Interventions on one platform are not necessarily limited to just that platform and can spill over to other platforms. For example, in May 2020 the social media platform Parler gained a significant number of new users, allegedly due to Twitter labeling President Trump's tweets as glorifying violence [8]. Therefore, it is no longer sufficient to study the impact of certain interventions in the context of one platform, but cross-platform research is necessary to understand the full impact of those interventions.

3 Conclusion

In recent years, fake news captured the attention of both the public and the research community. We now know considerably more about the scale and scope of fake news on social media and its distribution in the population. Building on these findings, this talk portrayed a path forward that calls for innovation in sampling techniques, a greater focus on detection tasks that aid fact-checkers, and methodology for evaluating intervention independently of social media platforms.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of Economic Perspectives **31**(2), 211–36 (2017)
2. Augenstein, I., Lioma, C., Wang, D., Lima, L.C., Hansen, C., Hansen, C., Simonsen, J.G.: MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking of Claims. arXiv preprint (Oct 2019), http://arxiv.org/abs/1909.03242
3. Dimensions.ai: Publications containing "fake news" in title or abstract in the years 2017-2021 (inclusive)., https://www.dimensions.ai/
4. Ecker, U.K., O'Reilly, Z., Reid, J.S., Chang, E.P.: The effectiveness of short-format refutational fact-checks. British Journal of Psychology **111**(1), 36–54 (2020), publisher: Wiley Online Library
5. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D.: Fake news on Twitter during the 2016 U.S. presidential election. Science **363**(6425), 374 (Jan 2019). https://doi.org/10.1126/science.aau2706
6. Guess, A., Nagler, J., Tucker, J.: Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science advances **5**(1) (2019). https://doi.org/10.1126/sciadv.aau4586, https://advances.sciencemag.org/content/5/1/eaau4586
7. Guess, A., Nyhan, B., Reifler, J.: Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. European Research Council **9**(3), 4 (2018)
8. Lewinski, J.S.: Parler App's #Twexit Movement Recruits Twitter Users In Wake Of Trump Social Media Wars. Forbes (May 2020), https://www.forbes.com/sites/johnscottlewinski/2020/05/29/parler-apps-twexit-movement-recruits-twitter-users-in-wake-of-trump-social-media-wars/
9. Linvill, D.L., Boatwright, B.C., Grant, W.J., Warren, P.L.: "THE RUSSIANS ARE HACKING MY BRAIN!" investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign. Computers in Human Behavior **99**, 292–300 (2019)
10. Nørregaard, J., Horne, B.D., Adali, S.: NELA-GT-2018. Harvard Dataverse (Jun 2019). https://doi.org/10.7910/DVN/ULHLCB
11. Pennycook, G., Bear, A., Collins, E.T., Rand, D.G.: The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. Management Science **66**(11), 4944–4957 (2020)
12. Pennycook, G., Rand, D.G.: The psychology of fake news. Trends in cognitive sciences (2021)
13. Salganik, M.J.: Asking quesetions. In: Bit by bit: Social research in the digital age. Princeton University Press (2019)
14. Shaar, S., Martino, G.D.S., Babulkov, N., Nakov, P.: That is a Known Lie: Detecting Previously Fact-Checked Claims. arXiv preprint (May 2020), http://arxiv.org/abs/2005.06058
15. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots. Nature communications **9**(1), 1–9 (2018)

16. Shu, K., Wang, S., Lee, D., Liu, H.: Disinformation, Misinformation, and Fake News in Social Media. Springer (2020)
17. Silverman, C.: This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed News (Nov 2016)
18. Wang, X., Yu, C., Baumgartner, S., Korn, F.: Relevant document discovery for fact-checking articles. In: Companion Proceedings of the The Web Conference 2018. p. 525–533. WWW '18, International World Wide Web Conferences Steering Committee (2018), https://doi.org/10.1145/3184558.3188723
19. Yaqub, W., Kakhidze, O., Brockman, M.L., Memon, N., Patil, S.: Effects of credibility indicators on social media news sharing intent. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–14 (2020)