Misinformation research, four years later

By Andy Guess

The shock of Donald Trump's unprecedented election corresponded to a flurry of activity among researchers, civil society groups, and foundations to understand what they missed about social media's role in fostering a degraded information ecosystem. The result was a surge of research about various forms of misinformation, which went hand in hand with increased scrutiny of social platforms.

Since that time, an extraordinary amount of effort went into preparations to avoid a repeat of the surprises of 2016. Meanwhile, the pandemic provided the additional challenge of combating rampant misinformation about the virus during a time of uncertainty around a constantly evolving global health threat. A somewhat surprising result of all this has been a willingness by platforms to employ increasingly aggressive tactics designed to reduce the spread of misleading content, such as claims of voter fraud and vaccine skepticism. At the same time, a number of anticipated threats — major foreign influence operations, deepfakes — did not seem to emerge.

Now that the election has passed and efforts to subvert the outcome have failed, where does all this leave misinformation researchers? Before offering some possible answers, I will begin with a quick overview of how I think scholars responded to newly salient questions about the role of online misinformation in political behavior over the past four years.

There were several strands in the emerging literature. First, researchers asked, what is the *prevalence* of online misinformation — who sees what, and when? Who shares fake news? There were basic descriptive questions about the extent to which fake news suffused the larger information environment in 2016 and beyond; whether exposure to misinformation and dissemination were related; and, at the individual level, whether people in certain groups or with certain traits or characteristics are more susceptible to online misinformation. These studies emphasize that fake news tends to reach a limited but highly polarized segment of the population, and that people are more likely to see and engage with congenial misinformation. Older people (especially age 65 and over) also appear to both encounter and share more online fake news.

Second, preexisting research literature examines the effectiveness of fact-checking on misperceptions adapted to the specific forms that misinformation takes on social media, as well as specific kinds of interventions used by platforms. These studies ask, are people receptive to factual information? What are the most effective ways of counteracting misinformation? Under which conditions do people resist corrections?

Generally speaking, these studies find that people update factual beliefs according to the information they are presented, even if this rarely changes attitudes about political figures or parties. Moreover, genuine instances of "backfire" — in which people resist corrective information to such an extent that it actually strengthens their prior misperceptions — appear to be rare. In the context of social media, this suggests that warning labels attached to content from fake news purveyors, and prominent notices about fact checks of questionable claims, are likely effective, though the magnitude is modest.

A lot of this work occurred in a kind of dialogue with researchers at the platforms who were concurrently developing and testing solutions of their own. This can be seen in the way that flags for disputed sources and the design of fact checks were sometimes justified by references to external scholarly research, which in turn has been inspired by the platforms' activities.

Third, researchers across social science disciplines have sought to explore the underpinnings of belief and sharing of fake news on social media. Looking beyond the specific circumstances of social media during a contentious election season in a polarized electorate, what cognitive or other tendencies underlie these phenomena? The answers differ somewhat depending on the outcome of interest (belief or sharing), but the list of suspects includes motivated reasoning driven by partisan animosity; tendencies toward cognitive reflection; digital literacy or skills correlated with age; and social influence.

Ultimately, these research strands have succeeded in providing descriptive and causal evidence on the scope of the misinformation problem and the kinds of relatively modest interventions that platforms could use to improve the quality of their feeds. In large part, the questions analysts focus on are a function of what is feasible in terms of data availability, research design, and ethics. As a result, there are plenty of concerns about the generalizability of this research (across platforms, countries, and time), as well as its scope. In particular, two critiques have been leveled at mainstream misinformation research. The first is that it fails to challenge the dominant business model of social platforms, which is premised on maximizing engagement. The second is that it often, but not always, abstracts away from the asymmetrically polarized political system and the larger partisan media ecosystem.

Meanwhile, platforms this year started rolling out efforts that haven't for the most part been the focus of existing research: banning ads, adding "frictions," reducing the reach of (or taking down entirely) misleading posts, and signal-boosting quality information around the elections and COVID-19. Although we lack reproducible evidence, these

efforts likely had a large impact. [From public reports](), it seems that even Facebook's relatively light-touch informational labels on false claims about the election reduced reshares of posts by 8%, while Twitter's nudge-like prompts resulted in a claimed 29% reduction in quote tweets of disputed claims. Even the 8% figure would be considered a large-effect size for most social science studies of interventions to reduce the spread of online misinformation. In other words, some of the most promising and aggressive approaches now being actively used by platforms — such as [downranking content via algorithms]() and adding frictions or nudges — haven't been systematically tested by independent researchers.

These developments should prompt reflection about the best way forward for research on misinformation. I'll focus on fact-checking research for now, since it is a prominent element of both social media companies' and news organizations' efforts to counteract misinformation, and I also continue to do work in this area.

Fact checks are an important part of the toolkit for confronting misinformation, and we should continue to advocate for their use by social media companies in partnership with professional fact-checking organizations. At the same time, we should acknowledge the limitations of this approach due to issues of scale and the lack of consistent ground truth. Ultimately, fact-checking is a mainstream journalistic practice that was never designed to solve platform-wide content moderation problems.

Furthermore, a surprising amount of what observers consider to be objectionable content isn't subject to factual verification. Take the claims of voter fraud surrounding the 2020 U.S. election. Before any assertions were challenged in court, they were literally unverifiable, meaning that content policies warning users about election misinformation were justified on other grounds. For example, in explaining its decision to remove the "Stop the Steal" group, Facebook referred not to online falsehoods or even encouraging violence but to "delegitimization of the election process."

A promising way forward for tractable research that could have outside policy impact on the way platforms operate is to increase focus on relatively low-effort nudges, primes, and skills modules that provide people with tools and competencies that can help them navigate their information feeds. These can range from priming people to focus on accuracy concerns to designing digital literacy interventions that have lasting effect. Despite their promise, however, there remains a large gap between claims about this kind of training and evidence about its effectiveness.

While nudges and primes are not in themselves a comprehensive policy prescription, I think it does suggest that misinformation research should avoid an excessive focus on

existing policy options, since we have seen that these can quickly change. A continually moving target means that the temporal validity of research on efforts to counteract misinformation may be low. One response is to double down on basic research of the kind that is already underway: understanding both the individual and structural factors combining to produce a state of information disorder. Some of these structural factors may be related to technological developments and the features of social platforms (such as algorithms). But not all of them will be, and this suggests that an integrated approach may be fruitful.

It may also be time to move beyond the dichotomies that we're used to — information/misinformation, fake/real, low-quality/high-quality — and instead ask how to affirm certain values. To start, some places could promote values such as democratic citizenship and healthy communities, but other values may appeal to different platforms in different parts of the world.

These ways forward are not without their own potential pitfalls. Doing basic research on the impact of social media and how to translate democratic values into concrete affordances and moderation approaches may increase the returns to collaboration with the platforms themselves, raising complex questions about privacy, ethics, and independence. Fortunately, the past four years have been a time of innovation on this front as well. A major challenge for the future will be to maintain the ability to think outside the box in terms of possible solutions, while also collaborating with private industry when possible in order to move forward our understanding of basic questions around the causes and consequences of misinformation, broadly speaking.

Andy Guess is an assistant professor of politics and public affairs at Princeton University.