Center for American Progress

GETTY IMAGES/JOHN LAMPARSKI

Fighting Coronavirus Misinformation and Disinformation

Preventive Product Recommendations for Social Media Platforms

By Erin Simpson and Adam Conner August 2020

Contents

Executive summary

Although online disinformation and misinformation about the coronavirus are different—the former is the intentional spreading of false or misleading information and the latter is the unintentional sharing of the same—both are a serious threat to public health. Social media platforms have facilitated an informational environment that, in combination with other factors, has complicated the public health response, enabled widespread confusion, and contributed to loss of life during the pandemic.

Looking ahead, the Center for American Progress expects disinformation and misinformation about the coronavirus to shift and worsen. As public health conditions vary more widely across the United States, this geographic variation will be an ideal vector for malicious actors to exploit. Without robust local media ecosystems, it will be especially difficult for social media platforms to moderate place-based disinformation and misinformation.

Long-term regulatory action will be needed to address the structural factors that contribute to an online environment in which misinformation and disinformation thrive. In the near term, social media platforms must do more to reduce the harm they facilitate, starting with fast-moving coronavirus misinformation and disinformation.

Social media platforms should go further in addressing coronavirus misinformation and disinformation by structurally altering how their websites function. For the sake of public health, social media platforms must change their product features designed to incentivize maximum engagement and amplify the most engaging posts over all others. Doing so will require fundamental changes to the user-facing products and to back-end algorithms that deliver content and make recommendations. Platforms must pair these changes with unprecedented transparency in order to enable independent researchers and civil society groups to appropriately study their effects.

With an eye toward proactively reducing mis/disinformation around the coronavirus crisis, this report makes specific recommendations of product changes that increase friction—anything that inhibits user action within a digital interface—and provide greater context. These principles are discussed in detail below, and suggestions are listed for convenience in the Appendix. Examples of changes recommended include:

- **Virality circuit breakers.** Platforms should detect, label, suspend algorithmic amplification, and prioritize rapid review and fact-checking of trending coronavirus content that displays reliable misinformation markers, which can be drawn from the existing body of coronavirus mis/disinformation.
- **Scan-and-suggest features.** Platforms should develop privacy-sensitive features to scan draft posts, detect drafts discussing the coronavirus, and suggest quality information to users or provide them cues about being thoughtful or aware of trending mis/disinformation prior to publication.
- **Subject matter context additions.** Social media platforms should embed quality information and relevant fact checks around posts on coronavirus topics. Providing in-post context by default can help equip users with the information they need to interpret the post content for themselves.

Overview

Disinformation online thrives in crisis.[1] Malicious actors capitalize on confusion, fear, and sorrow online for profit and political gain, intentionally spreading falsehoods and conspiracy and stoking engagement among social media users. Though a long-standing practice, this has perhaps never been more apparent than with COVID-19. Simultaneously, ongoing global attention and evolving scientific understanding of the novel coronavirus have created conditions for widespread sharing of misinformation—a problem in and of itself and a problem in the way it aids disinformation producers. Due to the prevalence of disinformation and misinformation on social media platforms, their use has become a health risk[2] during the coronavirus crisis. As the United States enters the next phase of the pandemic, which will bring even greater variation in public health conditions across the country, CAP is concerned that coronavirus disinformation and misinformation problems will intensify online.

Initial steps from social media platforms to mitigate COVID-19 mis/disinformation have sought to elevate "authoritative sources" of information, such as leading public health agencies, and to improve content moderation to reactively minimize harmful information.[3] Social media platforms' new commitments indicate a promising, if long overdue, shift. But as the United States faces the further disintegration of the shared reality of the pandemic, the need to do more than elevate authoritative information and reactively moderate harmful information is clear. Platforms can and must take additional, immediate actions to address coronavirus mis/disinformation by beginning to fundamentally alter their products, changing both the user-facing front ends and the algorithmically powered back ends that are invisible to the user.

The initial shared reality of the pandemic in the United States—with many states issuing stay-at-home orders to flatten the curve—has grown less universal. COVID-19 conditions have begun to vary substantially by geography: With some areas successfully suppressing the virus, some managing ongoing growth, and others grappling with re-outbreaks, one can expect a greater variety of state- or local-specific public health measures. The differentiation of public health conditions, lack of certainty around the coronavirus, and lack of local media resources are likely to lead to continued spread of misinformation. For malicious actors, the local variation in COVID-19 conditions and response is a ripe opportunity to sow division, cause chaos, and splinter what remains of a shared nationwide narrative and sense of reality. Due especially to the hollowing out of local media ecosystems over the past two decades,[4] communities are informationally vulnerable to place-based misinformation and disinformation. The coronavirus crisis presents an ideal vector to exploit. The issue of pandemic response is, as already demonstrated,[5] an easy informational vehicle for driving political polarization,[6] harassing others,[7] dehumanizing people of color,[8] and gaining power and attention online.[9]

There is a need for regulatory action against social media platforms. Effectively addressing online disinformation and misinformation problems will require regulatory change and structural reckoning with the fundamentally predatory elements of current business models. Ideally, this crisis will catalyze swift, systemic change from both Congress and companies, with many excellent proposals emerging along those lines from advocates and experts.[10] The systemic factors at play in terms of driving harmful content must also be considered, both on social media platforms—such as algorithmic optimization decisions[11]—and beyond—including systemic racism and coordinated behavior by white supremacist groups.[12] Regulatory action will be required to address these issues and other factors that create an online environment where mis/disinformation can thrive.

The problem the country faces is not a binary one of sorting through true or false information ... but instead a social, epistemic crisis fueled by changing media and political landscapes.

Until fundamental regulatory changes are made, however, there is an immediate need for mitigation around ongoing COVID-19 mis/disinformation. This report briefly explores the coming changes in the coronavirus crisis mis/disinformation landscape before making near-term recommendations to social media platforms on additional ways to mitigate harms. In this report, the authors wish to expand the conversation about COVID-19 response measures on social media to include proactive, near-term product strategies that could also mitigate pandemic disinformation. The authors recommend strategies to provide greater product context for users and to increase product friction; intentional friction would introduce front-end features that invite a user to be more thoughtful in engaging or sharing content, and back-end friction would alter the content algorithms that determine what content a user sees. These recommendations run counter to the frictionless product experience idealized by social media platforms, but they should serve as a call to action to others in reimagining the status quo.

Background

Misinformation about the coronavirus has plagued effective public health response from the start. Early on, the World Health Organization (WHO) warned of a so-called infodemic of online misinformation.[13] As public health officials urged the public to stay home and flatten the curve, experts raised the alarm that their messages were competing with a tide of misinformation and disinformation online.[14] The uncertainty surrounding the coronavirus, paired with intense global demand for information, created a perfect storm of speculation, conspiracy, and sharing of false or even harmful information. Complicating the situation, prominent public figures—including celebrities and politicians—were among the primary drivers of engagement around COVID-19 misinformation in early 2020.[15] For average users and influencers alike, coronavirus misinformation is so prevalent on social media that it has become increasingly difficult to avoid participating in spreading false or misleading narratives.

Compounding the problem in the United States, disinformation producers seized on the COVID-19 pandemic as a way to advance their goals and agendas by accelerating chaos online. Disinformation producers—and far-right and white supremacist ecosystems in particular[16]—have long sought to exploit crisis moments to amplify false, conspiratorial, and hateful narratives. Conspiracy theories have generally thrived in crisis,[17] but the modern social media environment and the sudden forced movement of attention online during stay-at-home orders have been a gift to malicious actors. Leveraging prevailing uncertainty, a demand for information, and an audience stuck online, these groups have effectively deployed disinformation strategies to pervert perceptions of public opinion and warp public discourse for their own gain. The coronavirus crisis was merely the most recent topic weaponized by the far-right disinformation ecosystem to spread racist narratives, undermine democratic institutions, and cause chaos.[18]

Defining disinformation and misinformation

Disinformation and misinformation differ on intent:[19] Disinformation is considered to be the intentional creation or sharing of false or misleading information, whereas spreading misinformation is considered to be unintentional sharing. Analytically, discerning intent requires understanding the creator's context and can be difficult to prove without identifying creators or uncovering coordination of efforts. Practically, the distinction can also be blurred in cases where a person doesn't care about the veracity of the information at all: Indeed, the spread of misinformation includes cases not only where people are fooled but also cases where people are enthusiastic about the content regardless of veracity because it supports their worldview[20] or where they are merely apathetic toward the truth. This apathy toward truth, or exhaustion with the difficulty in discerning it amid informational chaos online, is a feeling that disinformation producers have sought to increase. Russian information operations, for example, have long sought to erode trust in democratic institutions and processes by informationally exhausting Americans in hopes that they will give up or tune out.[21]

It's important to note that disinformation and misinformation are not always as straightforward as the sharing of outright lies: Conspiracy, misleading design or framing, opinion presented as fact, old facts presented as new facts, and the bent of each of these issues online toward prejudice and violent speech are also part of these problems.[22] While there are accounts and spaces dedicated to spreading mis/disinformation, these accounts often gain audiences by blending harmful information with legitimate news and innocuous posts or getting harmful messages amplified by more mainstream accounts. In this report, the authors use disinformation, misinformation, and their portmanteau (mis/disinformation) as synecdoche for representing the broader family of problems that contribute to informational chaos online.

Platforms, for their part, were quick to respond once the pandemic hit the United States. YouTube,[23] Facebook,[24] Instagram,[25] Reddit,[26] Twitter,[27] TikTok,[28] WhatsApp,[29] Snapchat,[30] and others pledged varying approaches to elevating authoritative information and stemming the tide of disinformation or coordinated inauthentic behavior, among other efforts. For discussion of actions taken thus far, experts at Public Knowledge,[31] WITNESS,[32] the EU DisinfoLab,[33] and New America's Open Technology Institute[34] have each put forth detailed analyses of platform actions around COVID-19. Platforms' enhanced efforts are needed urgently and are long overdue. Advocates and scholars have long called for varied but dramatic improvements to platform moderation systems.[35] Moreover, it is yet to be seen whether platforms will, as digital rights leaders from around the world have called for,[36] commit to data preservation and data sharing around COVID-19 content moderation to enable independent evaluation.

Even acknowledging the challenges in implementing increased content moderation enforcement during the pandemic, including unprecedented content moderation contractor office shutdowns and a shift to automated systems,[37] it has been sobering to see how little effect companies' unprecedented actions seem to have had on reducing the scale of the problem.[38] A recent fact sheet from the Reuters Institute for the Study of Journalism found that, "On Twitter, 59% of posts rated as false in our sample by fact-checkers remain up. On YouTube, 27% remain up, and on Facebook, 24% of false-rated content in our sample remains up without warning labels."[39] Researchers at Avaaz found that Facebook is "an epicentre of coronavirus misinformation," and came to similar conclusions: "Of the 41% of this misinformation content that remains on the platform without warning labels, 65% had been debunked by partners of Facebook's very own fact-checking program," and "over half (51%) of non-English misinformation content had no warning labels."[40] YouTube has taken unprecedented steps to elevate quality information about the coronavirus,[41] but highly politicized or conspiratorial health information receives more engagement relative to videos categorized as factual or neutral.[42] Further reports abound on instances of mislabeling, misapplying, slowly applying, or simply not having terms that functionally address harmful or false coronavirus content.[43]

The challenges that social media companies face in turning promises into results highlight a hard truth about tackling widespread mis/disinformation: The problem the country faces is not a binary one of sorting through true or false information about COVID-19 as it evolves, but instead a social, epistemic crisis fueled by changing media and political landscapes—an evolution in which social media platforms have played a central role.[44] Though the initial misperception of the public health crisis as apolitical was one enabler of social media platforms in their unprecedented response to mis/disinformation, enforcement and implementation of these changes have been complicated by the inherently political nature of amplifying information and specifically by the increasingly polarized reaction to the pandemic in the United States. Mis/disinformation from public officials and conservative media ecosystems[45] have accelerated this polarization and further complicated efforts to slow coronavirus mis/disinformation and the coronavirus itself.[46]

Helping people find trustworthy information about COVID-19 is also complicated by the dynamic social and scientific processes that are unfolding live: This virus is new, uncertainties abound, and scientific understanding is rapidly evolving.[47] Public institutions or official sources are themselves subject to error, manipulation, and politicization. When the Centers for Disease Control and Prevention (CDC) initially recommended against masks—for a mix of public health, political, and economic reasons—skepticism in public discourse raised the alarm that this may not be wise.[48] The CDC rightly reversed that recommendation[49]—though lingering concerns remain—and the WHO only recently made a recommendation.[50] Moreover, prominent public figures have both intentionally and unintentionally played a key role in spreading coronavirus mis/disinformation. In a recent analysis, researchers at the Reuters Institute found that "top-down misinformation from politicians, celebrities, and other prominent public figures made up just 20% of the claims in our sample but accounted for 69% of total social media engagement."[51] For example, transportation entrepreneur Elon Musk has provided a steady stream of misleading or false information about the coronavirus crisis to his large number of social media followers since the beginning of the crisis, including promoting unproven cures and repeatedly downplaying the public health threat.[52]

Unofficial, or interpersonal sources—discussion among friends and family about what's happening—also play an important role in every social circle for making sense of it all. Crisis informatics expert Kate Starbird terms this process "collective sensemaking," which is a key human response to crisis.[53] While the individual interpersonal sources used may not be "authoritative," group sensemaking is especially important in a time when the nation's journalistic institutions are hurting for funding, laying off workers, or shuttering—due in no small part to social media's capture of the online advertising market. Moreover, trust in government hovered at historic lows even before the recent protests against police brutality.[54]

The inherent challenges in uplifting quality information and dampening coronavirus mis/disinformation will grow more, not less, acute as the nation enters the next phase of the coronavirus crisis, which could last for months or even years. Substantial uncertainty remains about COVID-19, and the public will need to continue grappling with various aspects of pandemic response at regional, state, and local levels. But as these conditions begin to vary more widely over time, it will be increasingly difficult for platforms to have eyes on the sensemaking process playing out in their spaces. Platforms have struggled to keep pace with national guidance, which has been itself erratic at times, and this challenge will only compound with the growing variation in pandemic conditions. Such on-the-ground variation, combined with a demand for information and a lack of local media outlets, may also facilitate place-based misinformation.

COVID-19 is an ideal informational vehicle for disinformation producers to exploit for creating false perceptions of reality locally and elsewhere. With limited quality information sources at state and local levels, it is exceedingly easy to misrepresent local-level events and conditions to cause chaos and provide fake evidence or a sense of momentum for broader disinformation narratives. Effective public health responses could be derailed by place-based disinformation. As a consequence, public perception of the pandemic in the United States could be drastically influenced.

Coming shifts in the pandemic landscape will be difficult to navigate through reactive content moderation strategies alone, particularly at a time when platforms are ill-equipped to safely support the work of content moderators and to reorient their automated systems to tackle the emerging issues presented by COVID-19.[55] Platforms can and must do more to improve content moderation practices and algorithms, more effectively and consistently enforce their content policies, and improve workplace conditions for content moderators.[56]

Critically, the sheer scale of this challenge requires that every tool in the toolbox be on the table, especially the very products with which users interact. Instead of primarily being reactive to the spread of mis/disinformation, it is imperative that the very interfaces that help generate, incentivize, or amplify harmful coronavirus information be fully considered. Changes to the user-facing product must be explored in order to prevent the creation and spread of mis/disinformation along with algorithmic changes to prevent the spread of mis/disinformation once it is live. These product changes must be deployed and tested by the companies with a rapidity that matches the urgency of the moment. If paired with unprecedented transparency measures to help the public understand and to improve existing efforts, all of these are steps that can be implemented now, during the coronavirus crisis, and can be assessed as time goes on. There is a narrow window for platforms to make product changes to stymie the worst effects of this upcoming shift in COVID-19 misinformation and disinformation.

Efforts to uplift authoritative information and more effectively moderate harmful information posted to platforms should not obscure the fact that there is also a broader range of structural product features that social media companies have created and designed that could be altered to address the problem.

Recommendations for platforms

While the particular implementations will vary across the style and constraints of each platform, CAP believes that product-level changes focused on increased context and friction are necessary to proactively slow the tide of pandemic mis/disinformation. Product-level changes would adjust the content users see and the interfaces they use to interact with that content, such as changes to the interface on sharing a post or how the information in that post is presented. While the burden of managing mis/disinformation should not be solely shifted to users, CAP believes that design and features that help people navigate informational chaos are a necessary component of reducing the harms of disinformation in the near term. These recommendations are discussed in narrative form below and listed along with additional suggestions in the Appendix.

Aid users by introducing friction

Social media companies have long optimized for "frictionless" experiences. Within user experience design, friction is generally understood to be anything that inhibits user action within a digital interface,[57] particularly anything that requires an additional click or screen. Companies reduce friction to make it as easy as possible for users to engage and spend as long as possible on the platform. However, the downsides are clear when it comes to understanding the spread of mis/disinformation: The user experience is hyperoptimized for maximum engagement, which incentivizes user engagement with and consumption of harmful content. This means that in general, users have little context for their sharing and few cues that might encourage them to be thoughtful about doing so—an especially acute problem for coronavirus mis/disinformation.

At present, there are a number of product features that enable the creation and rapid spread of harmful misinformation. These features are working exactly as intended—making it easy to create and share content and then amplifying that content because it is highly engaging. These features, as part of a frictionless user experience, work in tandem with content recommendation algorithms—those that determine what users see in newsfeeds, trending sections, homepages, and various recommendations sections. Capturing more user time means more advertising can be sold. User behavioral data drawn from engagement means that advertising can be targeted more precisely, and thus be more profitable. On both counts, social media algorithms are optimized for engagement at the expense of any other value. Frictionless user experiences incentivize sharing, and engagement algorithmic systems amplify the most engaging and outrageous content.

It's past time for companies to do more than just pledge to be better at stopping mis/disinformation. Companies must rethink the product features and algorithmic optimization that they've designed to incentivize mis/disinformation in the first place. Introducing friction could preserve collective sensemaking while changing incentives, and in doing so, it could slow the spread of mis/disinformation.[58] Information policy expert Ellen P. Goodman, in her scholarship on "Digital Information Fidelity and Friction,"[59] writes that "communicative friction is a design feature to support cognitive autonomy." Paul Ohm and Jonathan Frankle frame design principles consistent with this idea as "desirable inefficiencies."[60]

Predicated on greater transparency, CAP recommends changes that add both front-end friction and back-end friction—that is to say, changes visible in the user interface as well as algorithmic changes that occur under the hood. In the context of coronavirus mis/disinformation, front-end changes would introduce cues that make users pause and think more about what they're sharing and with whom they're sharing. Back-end efforts in this spirit would include introducing internal structural measures to slow the creation and spread of potentially harmful content. While these changes would be intertwined and are mutually influential, both front-end and back-end approaches could discourage the creation of harmful content in the first place and help arrest its spread once published.

Back-end friction

Developing virality circuit breakers. Platforms have total control of algorithmic recommendation systems, but the opacity of platforms and their resistance to collaboration with researchers or regulators have complicated research in this area. In general, platforms have said that they're seeking to reduce the prevalence of various definitions of harmful COVID-19 content within their recommendation algorithms. Without eyes on those efforts, however, it's difficult to evaluate how they might affect the environment. Yet in reflecting on the possibilities of friction, CAP believes that there is an immediate algorithmic applicability in Goodman's discussion of parallels to circuit breakers in the financial market. Commenting on the adoption of friction in high-volatility algorithmic trading, Goodman writes: “The purpose of these circuit-breakers, in the view of the New York Stock Exchange, is to give investors ‘time to assimilate incoming information and the ability to make informed choices during periods of high market volatility.’ That is, it was expressly to create the space for the exercise of cognitive autonomy.”[61] Circuit breakers are considered important even for the relatively sophisticated and high-information actors in financial markets. The fact that the billions of social media users in a contextless, opaque online environment have no safety checks and fewer resources should give one pause.

Such a feature—a circuit breaker for viral mis/disinformation—has obvious potential for curbing harmful information about the coronavirus. Platforms will likely review harmful coronavirus information that reaches large audiences at some point, but reviewing and taking down posts after they have already gone viral often means the damage is already done.[62] Moreover, given that corrections struggle to garner the same levels of viewership as misinformation,[63] it’s difficult to mitigate the impact of harmful information once it’s out. Following Avaaz’s “Correcting the Record” research,[64] platforms should still pursue and invest in more robust methods of review; takedown; and direct, ad-hoc notification of users who interact with a harmful post with an appropriately designed fact check.[65] There may also be room to develop circuit breaker-like systems that arrest the spread of harmful viral content much earlier.

Platforms should identify trending coronavirus content that displays similar markers to already fact-checked coronavirus mis/disinformation; is trending within groups that incubate coronavirus mis/disinformation; or is posted by accounts that are serial publishers of coronavirus mis/disinformation. In order to create a body of work to inform the development of a viral circuit breaker, platforms should begin by using internal data from past user interactions and identified examples of viral COVID-19 misinformation to retroactively examine the spread of previous misinformation. That analysis should then be used to identify common patterns among viral COVID-19 disinformation to model the impact of potential interventions. Platforms should rapidly and transparently collaborate, test, and identify reliable indicators of harmful posts to carefully hone such a detection system—opening this process to contribution from researchers, journalists, technologists, and civil society groups across the world. Trending coronavirus posts that have reliable indicators of mis/disinformation should trigger rapid review by content moderation teams and get prioritization within fact-checking processes.

Until that review, fast-growing viral content believed to be related to the coronavirus could trigger an internal viral circuit breaker that temporarily prevents the content from algorithmic amplification in newsfeeds, appearing in trending topics, or other algorithmically aggregated and promoted avenues. Individual posting or message sharing could still occur—with the user warnings outlined below—but the algorithmic pause would allow the necessary time for a platform to review. Research will be needed to explore the holistic effects of such a pause, but it's possible that a short delay could prevent harms caused by the initial unchecked distribution of coronavirus misinformation and disinformation without overly punishing other coronavirus content. Such a feature may have prevented the viral spread of recent coronavirus conspiracy videos that rapidly republicized harmful, already debunked coronavirus falsehoods.[66]

For users, fast-growing viral content believed to be related to the coronavirus that is unchecked could cause a generic warning to pop up—such as "This content is spreading rapidly, but it has not yet been fact-checked"—until the content is able to be reviewed by platforms or third-party fact-checkers. Viral content should automatically be placed at the top of a queue for third-party fact-checking. There is, of course, reason for concern related to a "backfire effect"—placing a label on unchecked content could cast unnecessary doubt on an array of posts and potentially contribute to a user's overall apathy toward or exhaustion with the process of discerning the truth. Platforms should test numerous combinations of these interventions and conduct both short- and long-term polling and observation of the effects. Given the potential broad benefit for others, platforms should partner with researchers and enable independent study of this and other questions around such interventions.

While it would be difficult to develop a system to identify all harmful posts about the coronavirus as they begin trending, even flagging and reviewing some posts earlier could be an effective mitigation approach to these posts. It would also provide those users contributing to virality a chance to pause and reassess.

Rethinking autoplay. Platforms that enable video autoplay should rethink the algorithms behind video autoplay queues and suggested videos. Autoplay systems generate a queue of videos to watch next based on a viewer's browsing history and automatically start them following the conclusion of the viewer's current video. Video autoplay algorithms have been shown to be radicalizing forces for users.[67]

In the past, researchers have found that the YouTube video autoplay algorithm walks users toward increasingly extreme content.[68] This problem could be especially dangerous when it comes to spreading harmful coronavirus content. If platforms are unable to effectively prevent coronavirus mis/disinformation, they should retool video autoplay to play authoritative videos on the topic. YouTube,[69] TikTok,[70] and Snapchat[71] have all already curated authoritative coronavirus content. This authoritative content should be prioritized within next-to-play video queues on any subjects related to the coronavirus.

Adding friction for audience acquisition. Serial producers or sharers of coronavirus misinformation should be removed from recommendation algorithms for accounts to follow and friend and as groups to join. For groups or accounts that repeatedly violate terms around harmful coronavirus mis/disinformation, platforms should notify those accounts and group moderators of repeated violations and warn them of potential recommendation-algorithm docking or removal.

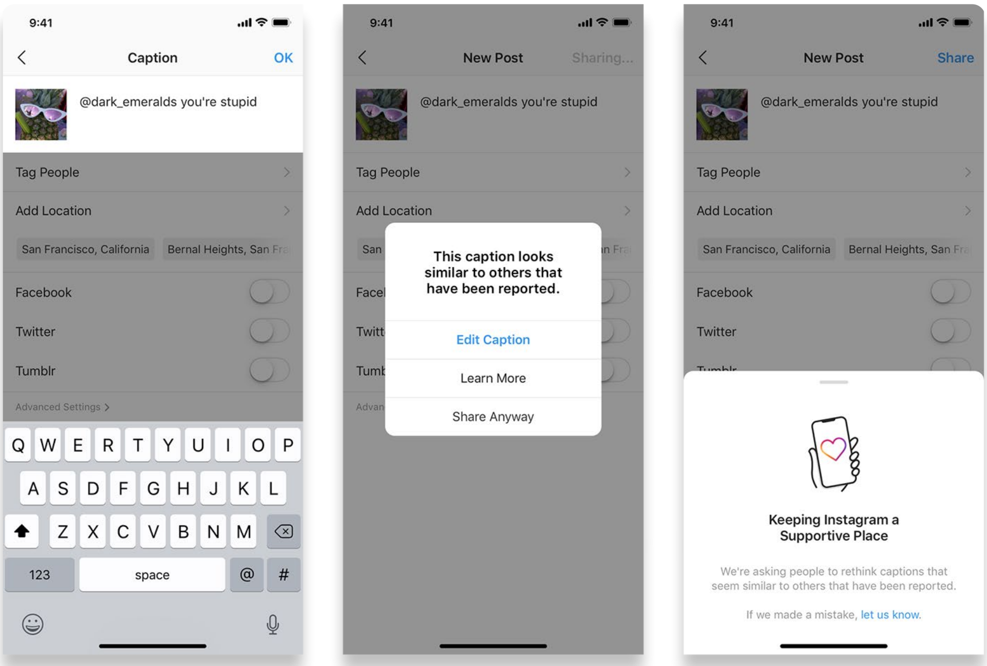
Adding friction for audience distribution. For accounts that repeatedly share false and harmful information about COVID-19 or COVID-19 response, social media platforms should push a warning to followers. Platform distribution algorithms should also take the sharing of content later found to be mis/disinformation into account in determining future distribution, notifying and docking future distribution if accounts have been shown to have a history of repeatedly spreading mis/disinformation.

If violations continue over time, existing members and followers should be forced to choose whether to stay/follow or leave/unfollow the sources of repeated mis/disinformation. For example, a prompt that forces followers to opt in to continue following could serve as a deterrent. This would change the default incentive to amass followers with assumed continued distribution and information consumption to an active choice that clearly alerts users that they're subscribing to accounts that are repeatedly sharing false information. Such a change would incentivize users who repeatedly share false or harmful information about COVID-19 to change their behavior or risk losing the audience that they have built up over time.

Front-end friction

Scan and suggest. Beneficial friction is already at work in mitigating harmful online content. In 2019, as a part of its anti-bullying efforts, Instagram started using artificial intelligence to scan draft captions and warn users if captions contained potentially offensive language similar to language that had been previously reported (see image below) and noted "promising" results.[72] Twitter has also started experimenting with prepublication revision tools.[73] The piloted feature being tested scans the draft text of a post, and if the draft includes language that's likely to be reported, users receive a prompt giving them the option to revise the reply before it's published. CAP encourages Instagram, Twitter, and every platform making parallel changes to publicly share more detailed findings on this test so that a broader community of experts can engage on what the holistic effects of such a change might be.

CAP proposes that a similar feature—scanning drafts and providing information or suggestive cues before publication—be brought to bear on the sharing of pandemic mis/disinformation. Platforms should pilot privacy-sensitive tools that detect keywords and cause an "Are you sure you're not spreading false information about COVID-19?" click-through screen to pop up when users are drafting a message or sharing content about a key COVID-19 topic such as social distancing, hand washing, vaccine information, or election modifications. For those uploading media, metatags or descriptions mentioning coronavirus key words could trigger similar pop-ups. In these type of actions, platforms could suggest relevant quality sources or fact checks or redirect users to dedicated information centers before a user posts, shares, forwards, or publishes content. Such an intervention could be deployed randomly as an occasional reminder for all users; deployed to all users around topics with trending disinformation and credible sources to provide; or deployed progressively more frequently for serial misinformation sharers. Research suggests that this approach may positively affect user behavior; it has been found that asking users to pause and consider the veracity of a headline slows the sharing of false information.[74] For users that do share false information, presenting them with a correction can dramatically reduce belief in false information.[75]



@dark_emeralds you're stupid

This caption looks similar to others that have been reported.

Keeping Instagram a Supportive Place

We're asking people to rethink captions that seem similar to others that have been reported.

If we made a mistake, let us know.

Credit: Instagram

Adding friction in messaging. Structural tweaks that add beneficial friction have already been shown to slow disinformation in messaging-centric platforms. Efforts from WhatsApp, a Facebook-owned messaging product, to stem viral disinformation have included not only a contextual "forwarded" icon but also user interface limitations on the number of simultaneous forwards and the number of groups one account can moderate.[76] After the global rollout of a limitation to restrict sharing on "highly forwarded messages" to only one chat, WhatsApp reported a "70% reduction in the number of highly forwarded messages sent on WhatsApp."[77] Reports suggest that Facebook Messenger, another Facebook-owned messaging product, has tested similar functionality but has not enabled it.[78] Given Messenger's reach and its integration with Facebook Groups, CAP believes a restriction on highly forwarded messages should be immediately activated for Messenger as well as other direct-messaging platforms, including Facebook-owned Instagram.

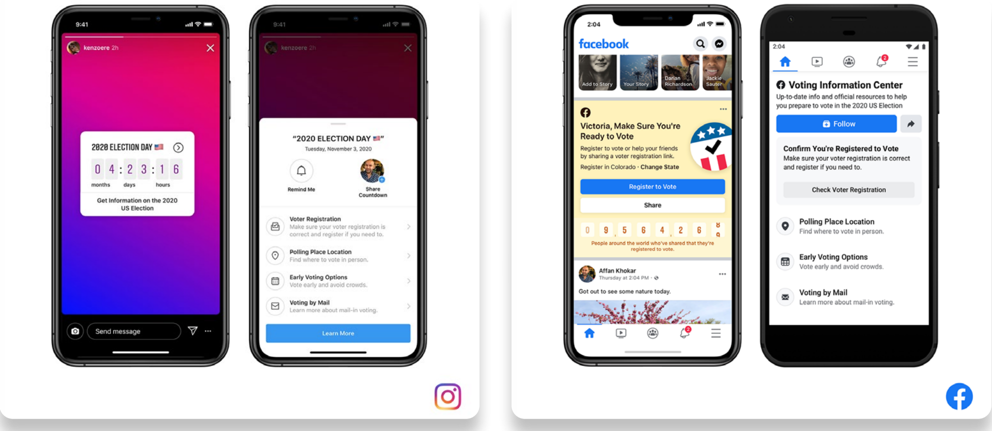
Context

Considering the ongoing challenges of moderating content well—which are only made worse by the uptick in traffic online[79]— finding other ways to aid users in processing and interpreting content must be a part of the mis/disinformation response. One strategy is to give users additional context to aid in processing information. Social media platforms tend to collapse social context for users,[80] displaying all content in a similar way despite varied sources, relationships, history, or purpose and intermediating relationships[81] between content producers and content consumers. As a result, these designs remove the social cues and contextual information that could help a user better interpret and make sense of what they're seeing. Recent research from scholars P.M. Krafft and Joan Donovan demonstrates that social media platform design tends to collate information into positive evidence collages; images are easily shared on social media, while textual conversations and overlays contesting those images are less easily shared.[82] When a user posts a false meme about the coronavirus, for example, other users might contest it in the comments and reactions. But if this meme is shared across platforms and, in some instances, on-platform, the false evidence in the image is carried over but the contesting text is lost; the context that would help new viewers be wary of the image is missing.

Platforms need not start from scratch. Already, efforts are being undertaken to provide context for users. With substantial thought toward how to maintain opportunity for new users and users without institutional affiliations, contextual information could greatly aid user discretion in processing and sharing coronavirus mis/disinformation. Contextual information could include information about the poster, the source of the content, and the subject matter of the content itself.

Subject matter context. Social media platforms should explore ways to provide context for posts relating to the coronavirus crisis. During the COVID-19 pandemic, YouTube has started displaying "fact-check panels" with content from eligible publishers above searches related to the coronavirus.[83] Snapchat is providing a filter series with vetted information on COVID-19.[84]

As a part of its coronavirus crisis response, Facebook created a COVID-19 Information Center to centralize and uplift authoritative sources on the coronavirus at the top of each newsfeed.[85] (see images below) Modeled on this approach, it later announced a Voting Information Center, which aims to curate credible information about voting, including posts from a user's local election authorities, registration guidance, vote-by-mail instructions, ID requirements, and more. (see images below) As part of the voting information effort, Facebook announced that it would automatically append a link to its Voting Information Center on all posts mentioning voting, regardless of content, veracity, or author.[86] Facebook will scan all posts and, if discussion of voting is detected, automatically append a link to the Voting Information Center on the post. Facebook should immediately take a similar scan-and-append approach to posts on coronavirus topics, curating credible and local coronavirus information and linking directly to the center on all posts mentioning the coronavirus. After the initial deployment, Facebook should aim to update its Voting Information Center and COVID-19 Information Center to provide more expanded personalized in-line information, without requiring a click for more information.



Make sure your voter registration is correct and register if you need to.

Find where to vote in person.

Vote early and avoid crowds.

Learn more about mail-in voting.

Victoria, Make Sure You're Ready to Vote

Register to vote or help your friends by sharing a voter registration link.

Register in Colorado · Change State

People around the world who've shared that they're registered to vote.

Got out to see some nature today.

Voting Information Center

Up-to-date info and official resources to help you prepare to vote in the 2020 US Election

Confirm You're Registered to Vote

Make sure your voter registration is correct and register if you need to.

Find where to vote in person.

Vote early and avoid crowds.

Learn more about mail-in voting.

Credit: Facebook

Other platforms should pursue similar efforts to embed quality information around posts on coronavirus topics for which the disinformation stakes are high. Providing in-post context can help equip users with the information that they need to interpret the post content for themselves. Given the difficulty of effective moderation under pandemic conditions, providing quality information on key subjects within user posts by default can help put the facts at users' fingertips.

Platforms should pair user content with labels, pop-ups, suggestions, or display of corresponding fact checks. In terms of ethical label design, as experts at First Draft note,[87] labels should be noticeable, consistent, nonconfrontational, and easy to process, as well as offer more information and not draw unnecessary attention to harmful content. In terms of video content, platforms should consider both user interface tweaks and short video warnings or contextual information that precede user content on sensitive topics.

The "truth sandwich" context. In addition to basic information on key areas of public interest, social media platforms should develop dynamic systems to automatically provide in-feed context on topics trending with disinformation. This could repurpose the existing ability of many platforms to deliver targeted contextual advertising and allow quality context to be provided before and after any potential disinformation. This builds on the "truth sandwich" concept suggested by psychologist George Lakoff to counter lies and disinformation, diminishing the harm of disinformation with authoritative sources.[88]

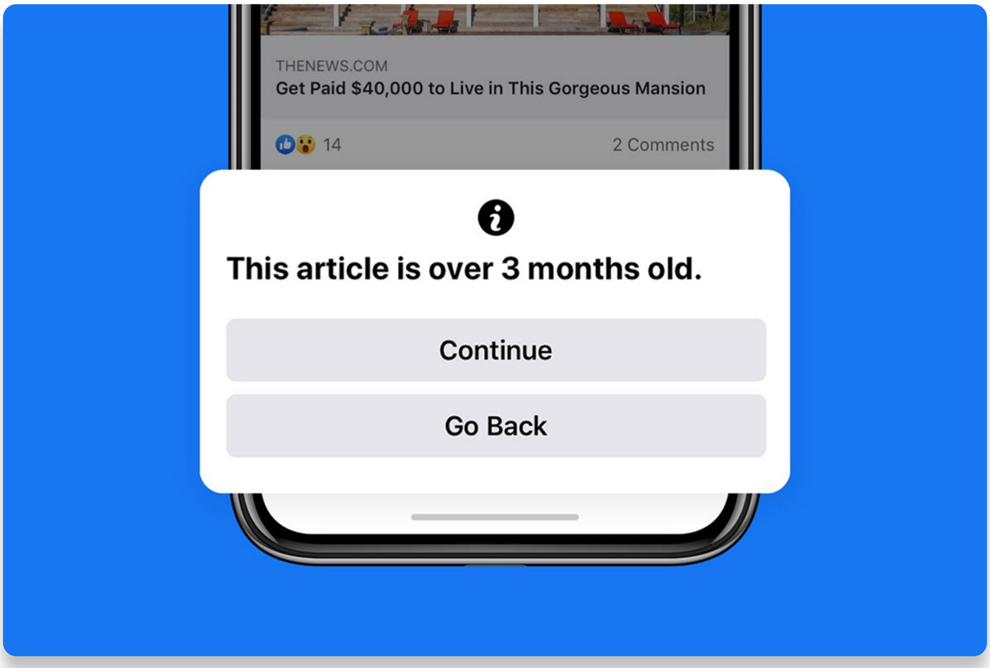
For example, the false conspiracy theory that 5G caused the coronavirus would have been a great candidate for trend-specific contextual fact checks. This conspiracy theory showed how mis/disinformation can rapidly spread out of control and cause real-world harms—in this case, harassment of telecom workers and arson attempts against cell towers in Western Europe.[89] Imagine that any posts with the words "5G" and "COVID-19"—those that did not violate terms or had not yet been flagged or reviewed—were surrounded above and below with authoritative information about COVID-19—the "truth sandwich." Discrete crisis events may merit general contextual information for on-topic posts as a way to help mitigate the spread of misinformation in the aftermath of a crisis.

Context for the poster. Contextual information should also include more detailed and accessible information about the poster within a post interface. Basic information such as location, relationship to the reader, duration on the platform, institutional affiliations, history on the platform, ad purchases on the platform, paid posts on the platform, and verified expertise in key areas of public interest could all help users weigh posts in different contexts. Twitter, for example, is rushing to verify more health sources to help elevate informed perspectives about the coronavirus.[90] Facebook is including location on posts by high-reach Pages and Instagram accounts.[91] These and similar changes would make it easier for users to process content.

In addition to basic information about the provenance, history, and relationship to the user, platforms should label accounts that repeatedly share false or misleading information about COVID-19. Platforms should go further in notifying accounts that publish and users who view or interact with harmfully inaccurate posts on COVID-19 about the specifics of their interaction, the offending content, its relationship to platform terms, and a relevant fact check.[92] This could happen either via notification or in context within existing feeds or streams. Research suggests that delivering specific corrections from fact-checkers could reduce belief in disinformation by half among social media users.[93]

Context on the source. Contextual information should also include information about the source of third-party content. At present, users have few cues to help them contextualize the source of third-party content posted by another user. Information on content publication dates, host sites, whether the URL is frequently fact-checked as false on the platform, and details inferred from top-level and second-level domains may help users catch red flags early. Drawing from on-platform details or verification program information about the third party could also help provide further information—for example, Twitter verifying candidates for elected office—as would linking out to off-platform groups with strong verification processes such as Wikipedia. Facebook has delivered animated contextual prompts for news stories since 2018, which includes drawing information from the source's Page and Wikipedia.[94] Facebook just announced a new click-through label for news stories that are more than 90 days old.[95] Instances such as these help users get basic contextual information at the moment of absorbing the headline message, rather than needing to go digging.

In addition to basic information about the source, platforms should go further in identifying and placing warning labels on sites that post content that is serially fact-checked as false or misleading. Platforms could also identify or place warning labels on look-alike media sites that falsely present themselves as legitimate journalistic outlets.[96]



THENEWS.COM

Get Paid $40,000 to Live in This Gorgeous Mansion

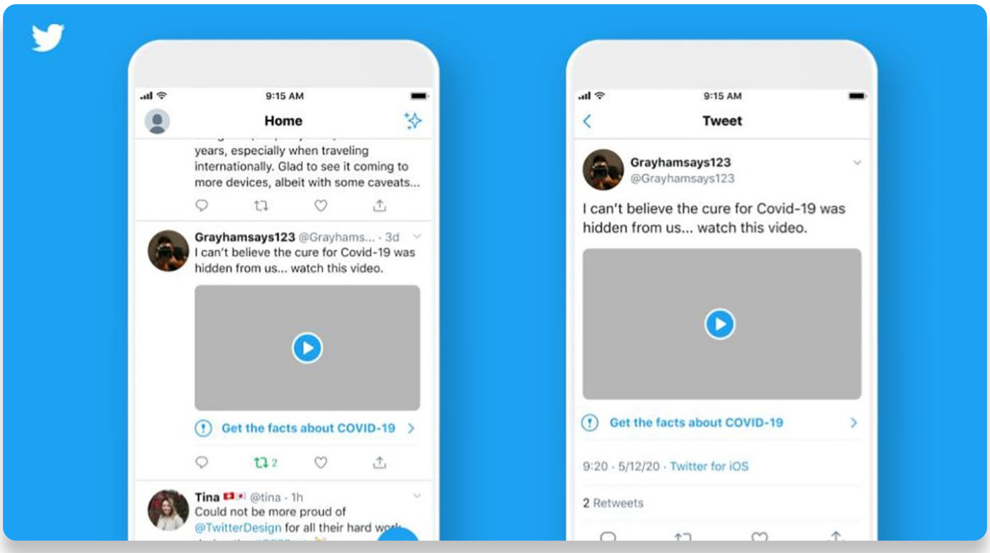
This article is over 3 months old.

Credit: Facebook

Cross- and off-platform context. Much of the content created on today's social media platforms is not just spread by the algorithms on a platform's website but offsite as well—as in off the originating platform and embedded, linked, or reposted somewhere else on the web. Often, the steps being taken to provide warnings or context occur only onsite and fail to transfer offsite. This ignores an important aspect of how information spreads online.

For example, the existing YouTube context bars do not appear for YouTube videos that appear in Google search.[97] Twitter's new labels on content (see image below) may not carry over when tweets are embedded offsite,[98] but the platform has noted there are updates coming in this area.

On-platform tweet, with context label:



years, especially when traveling internationally. Glad to see it coming to more devices, albeit with some caveats...

Grayhamsays123 @Grayhams... · 3d
I can't believe the cure for Covid-19 was hidden from us... watch this video.

Get the facts about COVID-19 >

2

Tina @tina · 1h
Could not be more proud of @TwitterDesign for all their hard work

Grayhamsays123
@Grayhamsays123

I can't believe the cure for Covid-19 was hidden from us... watch this video.

Get the facts about COVID-19 >

9:20 · 5/12/20 · Twitter for iOS

2 Retweets

Credit: Twitter

Platforms should commit to ensuring that steps they take are also carried through offsite, such as including warning labels or fact-check panels in embedded materials, requiring a click-through for materials deemed to have medical misinformation, and including additional context before or after suspected video misinformation. While it is true that platforms do not control the site that their content may be embedded on, they still often control what appears before or after that content—for example, with video advertising. Utilizing this existing advertising technology—for example, when a YouTube video that is suspected of medical misinformation is shared or embedded—could take advantage of YouTube's existing pre-roll video advertising infrastructure to show authoritative COVID-19 video information before and after the material. This would take the aforementioned "truth sandwich" concept to cross- or off-platform context.

Context in messaging. Coronavirus mis/disinformation that spreads within direct messaging platforms is difficult to arrest or provide with context. However, efforts such as WhatsApp's forwarded icon are a great example of small tweaks that provide context without needing to invade privacy.[99] The messages are still end-to-end encrypted, but WhatsApp provides the user with a double-arrow icon to let them know that a particular message has been forwarded more than five times and is thus unlikely to originate with a close contact. Knowing that disinformation spreads through messages on its app, WhatsApp has also sought to provide in-app context about the coronavirus through its Coronavirus Information Hub and through an integrated submission process to fact-checking organizations. Instagram and Facebook messaging features should adopt similar forwarded and highly forwarded labels.

Paying for context. Quality information isn't cheap to produce. While Facebook and Google have made contributions to fact-checking programs and local news support, it is not clear that financial support increases appropriately with the volume of work the companies may be generating. If social media giants are going to rely on Creative Commons work to serve context to users, those entities should be compensated for their work. Nonprofits such as Wikipedia have invested resources in creating quality informational processes for decades; platforms should appropriately contribute to Wikipedia, fact-checking organizations, and the journalism they rely on to help provide context for users. Even if their licensing structures allow for royalty-free use, platforms should commit to appropriate support and compensation.

Transparency

For all of these suggestions, however, an essential condition for taking swift action is unprecedented transparency. Addressing the public health threat that coronavirus mis/disinformation poses is a matter of public interest and merits public deliberation. But at present, the public is largely in the dark about what's been tried, what works, what hasn't, and how it's going. It is long past time for companies to deliver greater transparency[100] and increased data access[101] to researchers, regulators, and journalists, potentially via a tiered access disclosure system[102] that gives progressively higher levels of data access to researchers and regulators. As changes are being prepared, CAP urges platforms to heed the recommendations of numerous digital rights leaders to preserve data around COVID-19 content moderation.[103] Finally, efforts to mitigate mis/disinformation around the coronavirus could be greatly accelerated if past research was made public. Past internal studies and experimental data on efforts to address disinformation should be made public to aid in broader public understanding of disinformation and related issues.

Ethical, consistent moderation

Hopefully, this discussion is additive in terms of proactive product proposals for near-term mis/disinformation mitigation. While the authors did not seek to put forth recommendations directly addressing the challenges in content moderation itself, they echo others in restating that, while not the structural solution to these issues, content moderation will remain an essential part of addressing day-to-day mis/disinformation. If social media platforms do not have functional terms that address harmful mis/disinformation about the coronavirus crisis, those policies should be updated.

Moreover, platforms should apply rules transparently and consistently. To preserve freedom of expression, platforms must structurally incorporate human and civil rights considerations, adopt Change the Terms coalition standards against online hate,[104] and abide by the Santa Clara Principles, which outline standards for transparency, notice, and appeal around content moderation decisions.[105] Unfortunately, without greater transparency, it is very difficult to reconcile the contradictory findings of platforms and independent researchers around enforcement efforts. Therefore, platforms should both embrace the burden of proof to help the public better understand enforcement efforts and commit to greater transparency that enables researchers to independently explore them in depth.

Finally, platforms should not only invest in technical strategies that can better process context and at the same time be more sensitive to marginalized users, but they should also invest in the human content moderation workforce that is being asked to do a dangerous job under dangerous conditions. Platforms have long outsourced[106] the low-paid, technically challenging labor of sorting through traumatic, violent, and hateful material to a second class of workers who often aren't allowed to be publicly associated with the companies. Content moderators for major social media platforms should be treated as employees; their status as second-class contractors is a particularly disingenuous effort to hide the challenging and difficult work of content moderation. Facebook in particular—which just paid a $52 million settlement to content moderators who suffered from post-traumatic stress disorder and related conditions[107]—would do well to acknowledge their importance through representation on the Facebook Oversight Board, a commission Facebook created to comment on a selection of its moderation decisions. Social media platforms must also prioritize the health of those working in call center conditions, ensuring that systems are in place to prevent the spread of COVID-19 or making the privacy investments needed for moderators to work safely from home.

Conclusion

During a pandemic, disinformation and misinformation are a threat to public health. Bad actors have already sown confusion and division around the coronavirus and the public health response. As the United States enters the next phase of the COVID-19 pandemic and with public health conditions varying more widely among regions, states, and localities, coronavirus mis/disinformation is poised to intensify. Lasting solutions to mis/disinformation will require regulatory change that seeks a more open, competitive, and rights-respecting internet. In the near term, however, the set of tools for grappling with this crisis is extremely limited.

Beyond improved content moderation, CAP recommends that platforms explore product-level changes to provide context and increase beneficial friction as near-term, proactive mitigation methods for coronavirus mis/disinformation. The changes that flow from these recommendations will require significant company resources and enhanced transparency to ensure that they are implemented to curb false or harmful content about the pandemic and do not accidentally penalize the critical work of the press, public health organizations, advocates, and civil society during this time.

Admittedly, some of the recommendations and product suggestions may sound heavy handed to a generation of product managers and designers who have championed optimized and frictionless experiences—often to delightful or pro-social ends. But fear of intervention or choosing not to intervene is a design choice in and of itself. While the recommendations presented in this report may be somewhat draconian, the authors are confident that given the task of slowing coronavirus mis/disinformation and the permission to look holistically at freedom of expression, rather than only at engagement and ad revenue, more skillful, surgical, and appropriate interventions that incorporate context and friction can be found. The mis/disinformation crisis in this moment has laid bare the need for products to do more to prevent rather than primarily respond.

Finally, as platforms continue to grapple with the growing tide of COVID-19 mis/disinformation, they must publicly come to grips with the inescapably subjective, political nature of amplifying information for profit. CAP joins with others in calling for platforms to acknowledge their power, preserve sensemaking, increase space for user thoughtfulness, and be transparent about trade-offs and results of their efforts. Public health threats must be a matter of public deliberation. As companies continue efforts to mitigate pandemic mis/disinformation, CAP encourages them to work with the urgency and transparency that the moment requires.

About the authors

Erin Simpson is the associate director of Technology Policy at the Center for American Progress, where she's working toward an open, generative, rights-respecting internet and effective democratic regulatory infrastructure. Simpson served as the civil society lead for the Computational Propaganda Research Project at the University of Oxford, where she supported international civil and human rights leaders in preparing for disinformation and advocated for improved platform regulation in the European Union, United Kingdom, and United States. She was the founding director of programs at Civic Hall Labs and a Microsoft Civic Tech fellow. A Marshall scholar and Truman scholar, Simpson holds degrees from the University of Chicago and the Oxford Internet Institute at the University of Oxford.

Adam Conner is the vice president for Technology Policy at the Center. He leads the newly created Technology Policy team as its inaugural vice president with a focus on building a progressive technology policy platform and agenda. Conner has spent the past decade working at the intersection of technology, politics, policy, and elections as the first Washington, D.C., employee for several Silicon Valley companies. He was a spring 2018 resident fellow at the Harvard University Institute of Politics, where he led a study group titled "Platforms, Networks, and New Power Technology's Impact on Politics, Policy, and Elections," which focused on the rise of technology companies and their effect on politics and democracy.

Most recently, Conner was the first Washington employee for Slack Technologies, the fast-growing workplace communications startup, leading its engagement with federal, state, and local governments. Prior to that, Conner was vice president of Brigade, a civic engagement platform co-founded by Sean Parker. In 2007, Conner founded Facebook's Washington office. He spent seven years on the Facebook

Privacy and Public Policy team, where he created the company's government and political outreach efforts and directed the company's election efforts. Conner is a graduate of George Washington University's School of Media and Public Affairs and serves on the school's National Council. He is also on the board of the Roosevelt Institute. He hails from Los Alamos, New Mexico.

Appendix: Recommendations list

For convenience, a bulleted list of proposals for social media platforms set forth by this report is outlined below. As noted, not all recommendations apply consistently to all platforms, but CAP urges each to look carefully, ask how it may apply to them, and use their vast resources to test and improve upon these ideas. CAP further encourages platforms and others to propose other proactive interventions and product changes that create context and friction in order to reduce the harms of coronavirus mis/disinformation.

Each of the recommendations below has the potential to mitigate coronavirus mis/disinformation issues but also the potential for abuse or harm. Therefore, platforms must center human and civil rights best practices from the start in exploring these features and commit to unprecedented transparency measures that aid independent groups in their own evaluations. Platforms must work quickly to mitigate the harms caused by coronavirus mis/disinformation; public transparency around these changes is essential and will greatly accelerate the understanding of these problems.

Friction

Within user experience design, friction is generally understood to be anything that inhibits user action within a digital interface. Introducing beneficial friction into the individual user experience of information sharing and into the back-end amplification algorithms would be a way to slow mis/disinformation while preserving sensemaking.

Back-end friction

- Parallel to financial market circuit breakers,[108] platforms should develop virality circuit breakers. Trending coronavirus posts that have indicators of mis/disinformation should trigger rapid review by content moderation teams and get prioritization within fact-checking processes.

- Fast-growing coronavirus content should trigger an internal circuit breaker that temporarily prevents the content from algorithmic amplification in newsfeeds, appearing in trending topics, or other algorithmically aggregated and promoted avenues.

- Fast-growing coronavirus content that is unchecked should cause a generic warning to pop up, such as "This content is spreading rapidly, but it has not yet been fact-checked," until the content is able to be reviewed by platforms or third-party fact-checkers. Test multiple iterations, with short- and long-term effect observations, to ensure interventions are not generating unintended effects or causing backfire effects.

- Retool video autoplay queues to play only authoritative videos regarding the coronavirus.

- Serial producers or sharers of coronavirus mis/disinformation should be removed from recommendation algorithms for accounts to follow/friend and as groups to join.

- If violations continue over time for serial producers/sharers who are notified of coronavirus mis/disinformation, existing members or followers should be notified of repeated violations and forced to choose whether to stay/follow or leave/unfollow.

- Platform distribution algorithms should also take the sharing of content later found to be mis/disinformation into account in determining future distribution, notifying and docking future distribution for accounts that have shown to have a history of repeatedly spreading mis/disinformation.

Front-end friction

- Develop scan-and-suggest systems to proactively discourage coronavirus mis/disinformation. For draft, prepublication content that appears to violate terms around known areas of important or harmful coronavirus mis/disinformation, alert users of potential violations and ask if they'd like to revise their post before publication. (see Instagram caption alerts; Twitter reply revision experiment). Such a strategy could scan information in text-based posts, as well as captions for photo or video content.

- For draft content that appears to violate terms around known areas of important or harmful coronavirus mis/disinformation, alert users to credible fact-checking resources relevant to the topic. (see YouTube fact checking panels; Instagram caption alerts)

- For accounts that frequently distribute coronavirus misinformation, implement an "Are you sure you're not spreading false information about COVID-19?" click-through cue before a user can post, share, forward, or publish content. This intervention could appear progressively more frequently for accounts that continue to share mis/disinformation.

- For platforms with direct-messaging capabilities, limit the number of times a message can be forwarded simultaneously. (see WhatsApp's forwarding limitations)

Context

Giving users more cues and information to help contextualize coronavirus information and aid in processing coronavirus mis/disinformation, including context on the subject matter, the poster, and the third-party content sources.

- For posts on topics related to the coronavirus, automatically append links to information sources or dedicated information centers. (see Facebook's credible voting links on all posts about voting)
- For posts on topics related to coronavirus mis/disinformation, build in side-by-side displays of an appropriate fact check. (see YouTube's fact-check panels)
- Provide basic contextual information about the poster, including details such as location, relationship to the reader, duration on the platform, institutional affiliations, history on the platform, and verified expertise in key areas of public interest.
- Provide contextual information about the source of third-party content, such as content publication dates (see Facebook's old article prompts), host sites, details inferred from top-level and second-level domains, on-platform details or information from verification programs about the third party—for example, Twitter verifying candidates for elected office—whether the URL is frequently fact-checked on a platform, or off-platform information drawn from groups with strong verification processes such as Wikipedia.
- For in-feed content on topics trending with coronavirus mis/disinformation, take a "truth sandwich"[109] approach by pairing them with quality in-feed sources before and after potentially harmful posts.
- Label accounts that repeatedly share false or misleading information about COVID-19.
- Label posts that link to outside sites whose content is repeatedly fact-checked as false or misleading on a platform.
- Carry over any warning labels or fact checks when harmful content is shared across platforms.

- Provide contextual clues about how the content has been shared or promoted on-platform or across platforms.
- Label accounts of look-alike media sites that falsely present themselves as legitimate journalistic outlets.
- Notify publishing accounts who post COVID-19 mis/disinformation of the offending content and the relevant fact check.
- Notify users who view or interact with COVID-19 mis/disinformation about the specifics of their interaction, the offending content, its relationship to platform terms, and a relevant fact check.
- For high-reach accounts, platforms should provide information about the provenance, history, credentials and/or follower composition. (see Facebook Page transparency)
- For direct messages, provide forwarded labels on any forwarded messages. (see WhatsApp forwarded labels)
- For posts on coronavirus mis/disinformation that do not merit removal under terms or standards, provide labels on the post and the account, in addition to side-by-side fact-checking suggestions.
- For offsite video content—or all video content—include pre- or post-roll credible content.
- For videos on key coronavirus mis/disinformation topics, include a TV news-style ticker warning of frequent mis/disinformation on this topic and link out to relevant fact checks.
- For verified experts in key domains, provide domain-specific verification labels. (see Twitter's verification labels for candidates for public office, reimagined for domains with relevance to COVID-19)
- Major social media platforms should compensate any independent entities whose work is used to help provide quality, contextual information, including fact-checking organizations, Wikipedia, and independent media groups—even if licensing allows free use.

Endnotes

1 Kate Starbird, "How a Crisis Researcher Makes Sense of Covid-19 Misinformation," Medium, March 19, 2020, available at https://onezero.medium.com/reflecting-on-the-covid-19-infodemic-as-a-crisis-informatics-researcher-ce0656fa4d0a.

2 Daniel Allington and others, "Health-Protective Behaviour, Social Media Usage and Conspiracy Belief during the COVID-19 Public Health Emergency," *Psychological Medicine* (2020): 1–7, available at https://doi.org/10.1017/S003329172000224X.

3 EU DisinfoLab, "Platforms' Responses to COVID-19 Mis- and Disinformation," March 27, 2020, available at https://www.disinfo.eu/resources/covid-19/platforms-responses-to-covid-19-mis-and-disinformation/.

4 Matthew Hindman, "Less of the Same: The Lack of Local News on the Internet" (Washington: Federal Communications Commission, 2011), available at https://docs.fcc.gov/public/attachments/DOC-307476A1.pdf; Steven Waldman, "The Information Needs of Communities" (Washington: Working Group on Information Needs of Communities at the Federal Communications Commission, 2011), available at https://transition.fcc.gov/osp/inc-report/The_Information_Needs_of_Communities.pdf; Penelope Muse Abernathy, "The Expanding News Desert" (Chapel Hill, NC: UNC Center for Innovation and Sustainability in Local Media, 2018), available at https://www.usnewsdeserts.com/reports/expanding-news-desert/download-a-pdf-of-the-report/.

5 Melanie Smith, Erin McAweeney, and Léa Ronzaud, "The COVID-19 'Infodemic'" (New York: Graphika, 2020), available at https://public-assets.graphika.com/reports/Graphika_Report_Covid19_Infodemic.pdf.

6 Melissa Ryan, "Analysis: Against Excessive Quarantine Protests," Medium, April 20, 2020, available at https://medium.com/@melissaryan/analysis-against-excessive-quarantine-protests-fbe526b26097; Kathleen Hall Jamieson and Dolores Albarracín, "The Relation between Media Consumption and Misinformation at the Outset of the SARS-CoV-2 Pandemic in the US," *Harvard Kennedy School Misinformation Review* 1 (3) (2020): 1–22, available at https://doi.org/10.37016/mr-2020-012.

7 Berkman Klein Center, "Q&A: Renee DiResta on Disinformation and COVID-19," Medium, May 7, 2020, available at https://medium.com/berkman-klein-center/q-a-renee-diresta-on-disinformation-and-covid-19-7e285232d6e5.

8 Smith, McAweeney, and Ronzaud, "The COVID-19 'Infodemic'."

9 Cassie Miller, "White Supremacists See Coronavirus as an Opportunity," Southern Poverty Law Center, March 26, 2020, available at https://www.splcenter.org/hatewatch/2020/03/26/white-supremacists-see-coronavirus-opportunity.

10 Harold Feld, "The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms" (New York: Roosevelt Institute and Washington: Public Knowledge, 2019), available at https://www.publicknowledge.org/assets/uploads/documents/Case_for_the_Digital_Platform_Act_Harold_Feld_2019.pdf; Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, "Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online" (Philadelphia: Annenberg Public Policy Center, 2020), available at https://www.annenbergpublicpolicycenter.org/feature/transatlantic-working-group-freedom-and-accountability/; Ethan Zuckerman, "The Case for Digital Public Infrastructure" (New York: Knight First Amendment Institute at Columbia University, 2020), available at https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure; Nathalie Maréchal, Rebecca MacKinnon, and Jessica Dheere, "Getting to the Source of Infodemics: It's the Business Model" (Washington: New America Open Technology Institute, 2020), available at http://newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/; Stigler Committee on Digital Platforms, "Final Report" (Chicago: Stigler Center for the Study of the Economy and the State at the University of Chicago, 2019), available at https://research.chicagobooth.edu/stigler/media/news/committee-on-digital-platforms-final-report.

11 Maréchal, MacKinnon, and Dheere, "Getting to the Source of Infodemics: It's the Business Model."

12 Simon Clark, "How White Supremacy Returned to Mainstream Politics" (Washington: Center for American Progress, 2020), available at https://www.americanprogress.org/issues/security/reports/2020/07/01/482414/white-supremacy-returned-mainstream-politics/; Heidi Beirich and Wendy Via, "Generation Identity: International White Nationalist Movement Spreading on Twitter and YouTube," The Global Project Against Hate and Extremism, July 7, 2020, available at https://www.globalextremism.org/post/international-white-nationalist-movement-spreading-on-twitter-and-youtube.

13 World Health Organization, "Munich Security Conference Remarks: Transcript of Dr. Tedros Adhanom Ghebreyesus' speech at the World Health Organization, Munich Security Conference," February 15, 2020, available at https://www.who.int/dg/speeches/detail/munich-security-conference.

14 Brandy Zadrozny, "These disinformation researchers saw the coronavirus 'infodemic' coming," NBC News, May 14, 2020, available at https://www.nbcnews.com/tech/social-media/these-disinformation-researchers-saw-coronavirus-infodemic-coming-n1206911; Hubert Au, Philip N. Howard, and Jonathan Bright, "Coronavirus Misinformation: Weekly Briefings," Oxford Internet Institute, April 2020, available at https://comprop.oii.ox.ac.uk/research/coronavirus-weekly-briefings/; Brandi Collins-Dexter, "Canaries in the Coalmine: COVID-19 Misinformation and Black Communities" (Cambridge, MA: Shorenstein Center on Media, Politics and Public Policy at Harvard Kennedy School, 2020), available at https://shorensteincenter.org/canaries-in-the-coal-mine/; Institute for Strategic Dialogue Digital Research Unit, "Covid-19 Disinformation Briefing No.1" (London: 2020), available at https://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2020/03/Briefing-Covid-19.pdf.

15 Julie Ricard and Juliano Medeiros, "Using Misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil," *Harvard Kennedy School Misinformation Review* 1 (2) (2020): 1–6, available at https://misinforeview.hks.harvard.edu/article/using-misinformation-as-a-political-weapon-covid-19-and-bolsonaro-in-brazil/; J. Scott Brennen and others, "Types, Sources, and Claims of COVID-19 Misinformation," Reuters Institute for the Study of Journalism at the University of Oxford, April 7, 2020, available at https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation; Jim Waterson, "Influencers among 'key distributors' of coronavirus misinformation," *The Guardian*, April 8, 2020, available at https://www.theguardian.com/media/2020/apr/08/influencers-being-key-distributors-of-coronavirus-fake-news.

[16] Alice Marwick and Rebecca Lewis, "Media Manipulation and Disinformation Online" (New York: Data & Society, 2015), available at https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf; Miller, "White Supremacists See Coronavirus as an Opportunity"; Tech Transparency Project, "White Supremacist Groups Are Thriving on Facebook," Campaign for Accountability, May 21, 2020, available at https://www.techtransparencyproject.org/articles/white-supremacist-groups-are-thriving-on-facebook; Graphika, "Facebook's VDARE Takedown" (New York: 2020), available at https://public-assets.graphika.com/reports/graphika_report_vdare_takedown.pdf.

[17] Roger McNamee, "Social Media Platforms Claim Moderation Will Reduce Harassment, Disinformation and Conspiracies. It Won't," *Time*, June 24, 2020, available at https://time.com/5855733/social-media-platforms-claim-moderation-will-reduce-harassment-disinformation-and-conspiracies-it-wont/.

[18] Ryan, "Analysis: Against Excessive Quarantine Protests"; Carl Miller, "Far-right spreads Covid-19 'infodemic' on Facebook," BBC News, May 4, 2020, available at https://www.bbc.com/news/technology-52490430; Jason Wilson, "Disinformation and Scapegoating: How America's Far Right Is Responding to Coronavirus," *The Guardian*, March 19, 2020, available at https://www.theguardian.com/world/2020/mar/19/america-far-right-coronavirus-outbreak-trump-alex-jones; Mark Scott and Steven Overly, "Conspiracy theorists, far-right extremists around the world seize on the pandemic," *Politico,* May 13, 2020, available at https://www.politico.com/news/2020/05/12/trans-atlantic-conspiracy-coronavirus-251325.

[19] Caroline Jack, "Lexicon of Lies: Terms for Problematic Information" (New York: Data & Society, 2017), available at https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf.

[20] Alice E. Marwick, "Why Do People Share Fake News? A Sociotechnical Model of Media Effects," *Georgetown Law Technology Review* (474) (2018), available at https://georgetownlawtechreview.org/why-do-people-share-fake-news-a-sociotechnical-model-of-media-effects/GLTR-07-2018/.

[21] Adam B. Ellick and Adam Westbrook, "Operation Infektion: A Three-Part Video Series on Russian Disinformation," *The New York Times*, November 12, 2018, available at https://www.nytimes.com/2018/11/12/opinion/russia-meddling-disinformation-fake-news-elections.html; Sabrina Tavernise and Aidan Gardiner, "'No One Believes Anything': Voters Worn Out by a Fog of Political News," *The New York Times*, November 18, 2019, available at https://www.nytimes.com/2019/11/18/us/polls-media-fake-news.html.

[22] Claire Wardle, "Understanding Information Disorder" (New York: First Draft, 2019), available at https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x91514; Jack, "Lexicon of Lies: Terms for Problematic Information."

[23] YouTube, "Coronavirus Disease 2019 (COVID-19) Updates - YouTube Help," available at https://support.google.com/youtube/answer/9777243?hl=en (last accessed July 2020).

[24] Kang-Xing Jin, "Keeping People Safe and Informed About the Coronavirus," Facebook, July 16, 2020, available at https://about.fb.com/news/2020/07/coronavirus/.

[25] Ibid.

[26] u/worstnerd, "R/ModSupport - Misinformation and COVID-19: What Reddit Is Doing," Reddit, April 15, 2020, available at https://www.reddit.com/r/ModSupport/comments/g21ub7/misinformation_and_covid19_what_reddit_is_doing/.

[27] Twitter Inc., "Coronavirus: Staying Safe and Informed on Twitter," Twitter, April 3, 2020, available at https://blog.twitter.com/en_us/topics/company/2020/covid-19.html.

[28] TikTok, "Safety Center: Supporting Our Community Through COVID-19," available at https://www.tiktok.com/safety/resources/covid-19 (last accessed July 2020).

[29] WhatsApp, "How WhatsApp can help you stay connected during the coronavirus (COVID-19) pandemic," available at https://www.whatsapp.com/coronavirus/ (last accessed July 2020).

[30] Snap Inc., "Safety First," March 24, 2020, available at https://www.snap.com/en-US/news/post/safety-first.

[31] Lisa Macpherson, "How Are Platforms Responding to the Pandemic?", Public Knowledge, available at https://misinfotrackingreport.com/ (last accessed July 2020).

[32] Sam Gregory, Dia Kayyali, and Corin Faife, "COVID-19 Misinformation and Disinformation Responses: Sorting the Good from the Bad," WITNESS, available at https://blog.witness.org/2020/05/covid-19-misinformation-response-assessment/ (last accessed August 2020).

[33] EU DisinfoLab, "Platforms' Responses to COVID-19 Mis- and Disinformation."

[34] Spandana Singh and Koustubh "K.J." Bagchi, "How Internet Platforms Are Combating Disinformation and Misinformation in the Age of COVID-19," New America Open Technology Institute, available at http://newamerica.org/oti/reports/how-internet-platforms-are-combating-disinformation-and-misinformation-age-covid-19/ (last accessed August 2020).

[35] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018); Priyanjana Bengani, Mike Ananny, and Emily J. Bell, "Controlling the Conversation: The Ethics of Social Platforms and Content Moderation" (New York: Columbia University, 2018), available at https://doi.org/10.7916/D84F3751; Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven, CT: Yale University Press, 2019), available at https://www.jstor.org/stable/j.ctvhrcz0v; Change the Terms, "Reducing Hate Online," available at https://www.changetheterms.org/ (last accessed July 2020); David Kaye, "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression" (New York: U.N. General Assembly, 2018), available at https://freedex.org/wp-content/blogs.dir/2015/files/2018/05/G1809672.pdf.

[36] Emma Llansó, "COVID-19 Content Moderation Research Letter - in English, Spanish, & Arabic," Center for Democracy and Technology, April 22, 2020, available at https://cdt.org/insights/covid-19-content-moderation-research-letter/.

[37] Louise Matsakis and Paris Martineau, "Coronavirus Disrupts Social Media's First Line of Defense," *WIRED*, March 18, 2020, available at https://www.wired.com/story/coronavirus-social-media-automated-content-moderation/.

[38] Center for Countering Digital Hate, "#WilltoAct," available at https://www.counterhate.co.uk/willtoact (last accessed July 2020).

[39] Brennen and others, "Types, Sources, and Claims of COVID-19 Misinformation."

[40] Avaaz, "How Facebook Can Flatten the Curve of the Coronavirus Infodemic," April 15, 2020, available at https://fb.avaaz.org/campaign/en/facebook_coronavirus_misinformation/.

[41] YouTube, "Coronavirus Disease 2019 (COVID-19) Updates - YouTube Help."

42 Nahema Marchal, Hubert Au, and Philip N. Howard, "Coronavirus News and Information on YouTube," (Oxford, UK: Computational Propaganda Project at the Oxford Internet Institute, 2020), available at https://comprop.oii.ox.ac.uk/research/coronavirus-information-youtube/.

43 Queenie Wong, "More harm than good? Twitter struggles to label misleading COVID-19 tweets," CNET, May 25, 2020, available at https://www.cnet.com/news/more-harm-than-good-twitter-struggles-to-label-misleading-covid-19-tweets/; Kim Lyons, "Twitter Promises to Fine-Tune Its 5G Coronavirus Labeling after Unrelated Tweets Were Flagged," The Verge, June 27, 2020, available at https://www.theverge.com/2020/6/27/21305503/twitter-labels-5g-conspiracy-coronavirus; Sheera Frenkel, Ben Decker, and Davey Alba, "How the 'Plandemic' Movie and Its Falsehoods Spread Widely Online," *The New York Times*, May 20, 2020, available at https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html.

44 Sean Illing, "'Flood the Zone with Shit': How Misinformation Overwhelmed Our Democracy," Vox, January 16, 2020, available at https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-ban-non-misinformation; Robyn Caplan, "COVID-19 Misinformation Is a Crisis of Content Mediation," Brookings Institution TechStream, May 7, 2020, available at https://www.brookings.edu/techstream/covid-19-misinformation-is-a-crisis-of-content-mediation/.

45 Hall Jamieson and Albarracín, "The Relation between Media Consumption and Misinformation at the Outset of the SARS-CoV-2 Pandemic in the US"; Leonardo Bursztyn and others, "Misinformation During a Pandemic" (Chicago: Becker Friedman Institute, University of Chicago, 2020), available at https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_202044.pdf.

46 Christopher Ingraham, "New research explores how conservative media misinformation may have intensified the severity of the pandemic," *The Washington Post*, June 25, 2020, available at https://www.washingtonpost.com/business/2020/06/25/fox-news-hannity-coronavirus-misinformation/.

47 Siobhan Roberts, "Embracing the Uncertainties," *The New York Times*, April 7, 2020, available at https://www.nytimes.com/2020/04/07/science/coronavirus-uncertainty-scientific-trust.html.

48 Zeynep Tufekci, "Why Telling People They Don't Need Masks Backfired," *The New York Times*, March 17, 2020, available at https://www.nytimes.com/2020/03/17/opinion/coronavirus-face-masks.html.

49 Centers for Disease Control and Prevention, "Coronavirus Disease 2019 (COVID-19): Considerations for Wearing Masks," February 11, 2020, available at https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html.

50 BBC News, "Wear Masks in Public, WHO Says in New Advice," June 6, 2020, available at https://www.bbc.com/news/health-52945210.

51 Brennen and others, "Types, Sources, and Claims of COVID-19 Misinformation."

52 Russell Brandom, "Elon Musk is dangerously wrong about the novel coronavirus," The Verge, April 29, 2020, available at https://www.theverge.com/2020/4/29/21241180/elon-musk-coronavirus-conspiracy-misinformation-tesla.

53 Kate Starbird, "How to Cope with an Infodemic," Brookings Institution TechSteam, April 27, 2020, available at https://www.brookings.edu/techstream/how-to-cope-with-an-infodemic/.

54 Pew Research Center, "Public Trust in Government: 1958-2019," April 11, 2019, available at https://www.pewresearch.org/politics/2019/04/11/public-trust-in-government-1958-2019/.

55 Matsakis and Martineau, "Coronavirus Disrupts Social Media's First Line of Defense."

56 Elizabeth Dwoskin, Jeanne Whalen, and Regine Cabato, "Content moderators at YouTube, Facebook and Twitter see the worst of the web — and suffer silently," *The Washington Post*, July 25, 2019, available at https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/; Casey Newton, "Facebook Will Pay $52 Million in Settlement with Moderators Who Developed PTSD on the Job," The Verge, May 12, 2020, available at https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health.

57 Victoria Young, "Strategic UX: The Art of Reducing Friction," Telepathy, available at https://www.dtelepathy.com/blog/business/strategic-ux-the-art-of-reducing-friction (last accessed August 2020).

58 Lisa Fazio, "Pausing to consider why a headline is true or false can help reduce the sharing of false news," *Harvard Kennedy School Misinformation Review* 1 (2) (2020): 1–8, available at https://doi.org/10.37016/mr-2020-009; Justin Kosslyn, "The Internet Needs More Friction," Vice Motherboard blog, November 16, 2018, available at https://www.vice.com/en_us/article/3k9q33/the-internet-needs-more-friction; Zeynep Tufekci, @zeynep, January 21, 2019, 7:47 p.m. EST, Twitter, available at https://twitter.com/zeynep/status/1087511998299029504; Alexander B. Howard, "How Adding Friction To Group Messaging Can Help Defuse Disinformation," Defusing Disinfo, January 26, 2019, available at https://defusingdis.info/2019/01/26/how-adding-friction-to-group-messaging-can-help-defuse-disinformation/.

59 Ellen P. Goodman, "Digital Information Fidelity and Friction" (New York: Knight First Amendment Institute at Columbia University, 2020), available at https://knightcolumbia.org/content/digital-fidelity-and-friction.

60 Paul Ohm and Jonathan Frankle, "Desirable Inefficiency," *Florida Law Review* 70 (4) (2019): 777.

61 Donald Bernhardt and Marshall Eckblad, "Stock Market Crash of 1987," Federal Reserve History, November 22, 2013, available at https://www.federalreservehistory.org/essays/stock_market_crash_of_1987; Goodman, "Digital Information Fidelity and Friction."

62 Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359 (6380) (2018): 1146–1151, available at https://doi.org/10.1126/science.aap9559; Brendan Nyhan and Jason Reifler, "Misinformation and Fact-checking: Research Findings from Social Science" (Washington: New America Foundation, 2012), p. 28, available at https://www.dartmouth.edu/~nyhan/Misinformation_and_Fact-checking.pdf.

63 Michelle Ma, "Hold that RT: Much misinformation tweeted after 2013 Boston Marathon bombing," UW News, March 17, 2014, available at https://www.washington.edu/news/2014/03/17/hold-that-rt-much-misinformation-tweeted-after-2013-boston-marathon-bombing/.

64 Avaaz, "Correcting the Record" (London: Avaaz, 2020), available at https://fb.avaaz.org/campaign/en/correct_the_record_study/.

65 Vosoughi, Roy, and Aral, "The Spread of True and False News Online," pp. 1146–1151; Nyhan and Reifler, "Misinformation and Fact-Checking," p. 28.

66 Nick Statt, “Facebook Says Removing Viral COVID-19 Misinformation Video ‘Took Longer than It Should Have,’” The Verge, July 28, 2020, available at https://www.theverge.com/2020/7/28/21345674/facebook-covid-19-misinformation-breitbart-news-video-removal-response.

67 Rebecca Lewis, “Alternative Influence: Broadcasting the Reactionary Right on YouTube” (New York: Data & Society, 2018), available at https://datasociety.net/library/alternative-influence/; Zeynep Tufekci, “YouTube, the Great Radicalizer,” *The New York Times*, March 10, 2018, available at https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

68 Manoel Horta Ribeiro and others, “Auditing Radicalization Pathways on YouTube,” ArXiv:1908.08313 [Cs], December 4, 2019, available at http://arxiv.org/abs/1908.08313.

69 YouTube, “Coronavirus Disease 2019 (COVID-19) Updates - YouTube Help.”

70 TikTok, “Safety Center: Supporting Our Community Through COVID-19.”

71 Snap Inc, “Safety First.”

72 Instagram, “Our Progress on Leading the Fight Against Online Bullying,” Instagram Blog, available at https://about.instagram.com/blog/announcements/our-progress-on-leading-the-fight-against-online-bullying (last accessed July 2020).

73 Twitter Support, @TwitterSupport, May 5, 2020, 1:01 p.m. EST, Twitter, available at https://twitter.com/TwitterSupport/status/1257717113705414658/.

74 Fazio, “Pausing to Consider Why a Headline Is True or False Can Help Reduce the Sharing of False News.”

75 Avaaz, “Correcting the Record.”

76 WhatsApp Blog, “Keeping WhatsApp Personal and Private,” April 7, 2020, available at https://blog.whatsapp.com/Keeping-WhatsApp-Personal-and-Private.

77 Manish Singh, “WhatsApp’s New Limit Cuts Virality of ‘Highly Forwarded’ Messages by 70%,” TechCrunch, April 27, 2020, available at https://social.techcrunch.com/2020/04/27/whatsapps-new-limit-cuts-virality-of-highly-forwarded-messages-by-70/.

78 Jane Manchun Wong, @wongmjane, March 21, 2020, 7:06 a.m. EDT, Twitter, available at https://twitter.com/wongmjane/status/1241320233471623168.

79 Ella Koeze and Nathaniel Popper, “The Virus Changed the Way We Internet,” *The New York Times*, April 7, 2020, available at https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html.

80 Alice E. Marwick and danah boyd, “I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience,” *New Media & Society* 13 (1) (2011): 114–133, available at https://doi.org/10.1177/1461444810365313.

81 Ben Thompson, “Defining Aggregators,” Stratechery, September 26, 2017, available at https://stratechery.com/2017/defining-aggregators/.

82 P. M. Krafft and Joan Donovan, “Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign,” *Political Communication* 37 (2) (2020): 194–214, available at https://doi.org/10.1080/10584609.2019.1686094.

83 YouTube, “Coronavirus Disease 2019 (COVID-19) Updates - YouTube Help.”

84 Snap Inc., “Safety First.”

85 Jin, “Keeping People Safe and Informed About the Coronavirus - About Facebook.”

86 Mark Zuckerberg, Facebook, June 26, 2020, available at https://www.facebook.com/zuck/posts/10112048980882521.

87 Emily Saltz and others, “It matters how platforms label manipulated media. Here are 12 principles designers should follow,” First Draft, June 10, 2020, available at https://firstdraftnews.org:443/latest/it-matters-how-platforms-label-manipulated-media-here-are-12-principles-designers-should-follow/.

88 George Lakoff, @GeorgeLakoff, 10:37 a.m. ET, December 1, 2018, Twitter, available at https://twitter.com/georgelakoff/status/1068891959882846208?lang=en.

89 Ellen Nakashima, “DHS to advise telecom firms on preventing 5G cell tower attacks linked to coronavirus conspiracy theories,” *The Washington Post*, May 13, 2020, available at https://www.washingtonpost.com/national-security/dhs-to-advise-telecom-firms-on-preventing-5g-cell-tower-attacks-linked-to-coronavirus-conspiracy-theories/2020/05/13/6aa9eaa6-951f-11ea-82b4-c8db161ff6e5_story.html.

90 Twitter Inc., “Coronavirus: Staying Safe and Informed on Twitter.”

91 Anita Joseph and Georgina Sheedy-Collier, “Making Pages and Accounts More Transparent,” Facebook, April 22, 2020, available at https://about.fb.com/news/2020/04/page-and-account-transparency/.

92 Avaaz, “Correcting the Record.”

93 Ibid.

94 Jeff Smith, Alex Leavitt, and Grace Jackson, “Designing New Ways to Give Context to News Stories,” Facebook, April 8, 2018, available at https://about.fb.com/news/2018/04/inside-feed-article-context/.

95 John Hegeman, “Providing People With Additional Context About Content They Share,” Facebook, June 25, 2020, available at https://about.fb.com/news/2020/06/more-context-for-news-articles-and-other-content/.

96 Eric Lubbers, “There is no such thing as the Denver Guardian, despite that Facebook post you saw,” *The Denver Post*, November 7, 2016, available at https://www.denverpost.com/2016/11/05/there-is-no-such-thing-as-the-denver-guardian/.

97 YouTube Official Blog, “Expanding fact checks on YouTube to the United States,” April 28, 2020, available at https://youtube.googleblog.com/2020/04/expanding-fact-checks-on-youtube-to-united-states.html.

98 Yoel Roth and Nick Pickles, “Updating our approach to misleading information,” Twitter blog, May 11, 2020, available at https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html.

99 WhatsApp, “About forwarding limits,” available at https://faq.whatsapp.com/general/coronavirus-product-changes/about-forwarding-limits (last accessed July 2020).

100 Avaaz, “Legislative Principles for Tackling Disinformation,” available at https://fb.avaaz.org/campaign/en/disinfo_legislative_principles/ (last accessed July 2020); Kaye, “Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression.”

101 Llansó, “COVID-19 Content Moderation Research Letter - in English, Spanish, & Arabic.”

102 Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, "Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online."

103 Llansó, "COVID-19 Content Moderation Research Letter - in English, Spanish, & Arabic."

104 Change the Terms, "Reducing Hate Online."

105 SantaClaraPrinciples.org, "The Santa Clara Principles on Transparency & Accountability in Content Moderation," available at https://santaclaraprinciples.org/images/scp-og.png (last accessed July 2020).

106 Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media*.

107 Newton, "Facebook Will Pay $52 Million in Settlement with Moderators Who Developed PTSD on the Job."

108 Goodman, "Digital Information Fidelity and Friction."

109 Lakoff, Twitter.

Our Mission

The Center for American Progress is an independent, nonpartisan policy institute that is dedicated to improving the lives of all Americans, through bold, progressive ideas, as well as strong leadership and concerted action. Our aim is not just to change the conversation, but to change the country.

Our Values

As progressives, we believe America should be a land of boundless opportunity, where people can climb the ladder of economic mobility. We believe we owe it to future generations to protect the planet and promote peace and shared global prosperity.

And we believe an effective government can earn the trust of the American people, champion the common good over narrow self-interest, and harness the strength of our diversity.

Our Approach

We develop new policy ideas, challenge the media to cover the issues that truly matter, and shape the national debate. With policy teams in major issue areas, American Progress can think creatively at the cross-section of traditional boundaries to develop ideas for policymakers that lead to real change. By employing an extensive communications and outreach effort that we adapt to a rapidly changing media landscape, we move our ideas aggressively in the national policy debate.

Center for American Progress

1333 H STREET, NW, 10TH FLOOR, WASHINGTON, DC 20005 • TEL: 202-682-1611 • FAX: 202-682-1867 • WWW.AMERICANPROGRESS.ORG