In the lead-up to the 2020 US election, platforms have been taking an increasingly proactive approach to addressing potential risks posed by electoral misinformation, attempts at voter suppression or intimidation, and other electoral issues occurring on their sites. The period around elections poses a unique set of risks and challenges for platforms in moderating content. For one thing, the stakes of getting things wrong are unusually high: removing legitimate content or failing to remove misleading content could have consequences for the democratic process itself. These concerns arise around coordinated inauthentic activity from both domestic and foreign actors, as well as the ability for public figures with power to illegitimately use platforms to undermine the electoral process for personal gain. There are also particular forms of illegality which only arise around elections, such as voter intimidation. Because of concerns raised about the role of platforms during the 2016 US election, there has been particular attention paid and effort expended on how platforms approach electoral content in the lead up to the 2020 US election. The lessons learned from this context, however, are likely to have implications for how platforms approach elections globally in the future. It is therefore important to understand what steps platforms have taken, how they potentially impact human rights, and how rights like freedom of expression are balanced against the need to protect the democratic process and the right to free and fair elections.
Evolving Policies on Election Misinformation and Integrity of Civic Discourse

Ahead of the 2020 US election, the majority of platforms have established dedicated policies that identify what categories of election-related content will be subject to platform action. Until fairly recently these policies have focused on regulating specific misleading claims that could affect voter participation and/or lead to voter suppression. However, in recent months and weeks, the platforms have expanded their policies to cover additional categories of content that are more ambiguous including non-specific claims that could impact the election more broadly. For instance, in mid-September, Twitter broadened the scope of its civic integrity policy to include “false or misleading information intended to undermine public confidence” in the election - signalling a policy shift to address content that could create general confusion around the election. Facebook similarly expanded its voter interference policy to include “implicit misrepresentations about voting” in addition to “explicit misrepresentations.” This evolution in platform policies reflects an emerging trade-off from a human rights perspective. As platforms take on greater responsibility for addressing a wider range of content that could adversely impact the election, they will also be required to exercise more subjective judgment to decide what content they will act on. Such decisions, impacting on people’s ability to exercise their rights to free expression, are being made by the platforms as private actors who are unaccountable to the public and whose decision making processes remain opaque.

This electoral period also seems to have accelerated an ongoing shift at platforms away from the binary of leaving content up or taking it down, and towards a wider range of approaches including the application of labels that provide context or fact check misleading information. Facebook and Twitter have both updated their policies around elections to include the use of labels for cases of misleading information that may not meet the standard for removal and to provide additional context for certain types of content. Moderation approaches which move beyond removal provide platforms with better tools to balance the desire to protect free expression with the need to ensure that platforms are not used to disrupt democratic processes.

Changes in How Public Figures are Moderated

One of the major debates that has come to the forefront during the 2020 US election period is how platforms will deal with content from public officials that violate their policies. Both Twitter and Facebook have policies that provide exemptions to content moderation for elected officials and candidates, because of the public interest in voters knowing what their leaders are saying. Facebook does not subject posts from politicians to fact checking, for example, and both platforms will leave posts that would otherwise violate their standards up if they are posted by politicians.

In the lead-up to the election (and also motivated by posts by President Donald Trump and other world leaders that contained misinformation about the novel coronavirus) Twitter clarified its public interest exception and provided details about when posts from politicians would be removed or labeled, emphasizing that the likelihood of real world harm will guide decisions about how strong moderation should be against content from public figures. Several platforms (e.g., Facebook) also clarified that these exceptions do not apply to platform enforcement of their specific election-related misinformation policies ahead of the 2020 election. In practice, Twitter has taken a much more proactive role in moderating the speech of public officials, particularly President Trump. The company labeled one of Trump’s tweets for the first time in May, placing a label encouraging users to get the facts on mail-in voting.

Yet, enforcement of election-related misinformation policies remains significantly uneven against political figures. There have been several instances where Facebook and Twitter have declined to take action against misleading claims about the electoral process shared by President Trump, even when such content would appear to fall into categories they have said are actionable for political figures. This unevenness in enforcement is an additional barrier to users understanding the political landscape, and it creates opportunities for powerful individuals to misuse platforms to their advantage. To the extent that public interest exemptions exist, they should be clearly and narrowly defined, and enforcement should be consistent.
Changes to Ad Policies in the Electoral Period

Beyond organic content on the platform, political advertising and the rules that govern it also have critical implications during electoral periods. Different platforms take dramatically different approaches to political ads. Both Twitter and TikTok banned political ads on their platforms altogether in the fall of 2019. This move was celebrated in some quarters, while others worried that this eliminated a major avenue for smaller political candidates and groups to get their message across. Google has established specific limits on how political ads can be targeted towards users, and clarified that all ads, including political ads, are prohibited from making false claims. Facebook has continued to defend its largely hands-off approach to regulating political ads (including explicitly declining to fact-check ads).

However, in the immediate lead up to the 2020 elections, many of the platforms have converged on announcing temporary, time-bound bans on political ads in recognition of the heightened impact that misleading political ads could have in the immediate periods before and after election day. Facebook has announced that it will prohibit political ads in the week prior to the election. Their motivation is similar to that of countries that institute media blackouts in advance of elections: too close to an election, there is not sufficient time for candidates to respond to negative advertising or false claims. This is particularly important given that Facebook does not fact check political ads. Facebook and Google have also both announced a ban on ads for an undisclosed period after the election if the results are still unclear. This is in response to concerns that amid the increased uncertainty caused by additional mail-in voting, a candidate might try to use advertising to declare victory before all votes are finalized.

Structural Tools to Limit the Virality of Potentially Harmful Election Content

Platforms have also started to take actions to limit the visibility of potentially misleading content by tweaking aspects of platform architecture to reduce the likelihood that election-related misinformation will be amplified. Most recently, Twitter announced a series of actions that it will take during the pre-election period designed to limit the virality of misleading information. These actions include new prompts to nudge users to “quote tweet” rather than simply “retweet” content (i.e., adding commentary to retweeted content) and removing recommendations of content posted by accounts that a user is not already following. This is in addition to existing policies which limit algorithmic promotion of content that has been labeled as being misleading.

Proactive Provision of Authoritative Information

Several of the platforms have also developed new features to proactively provide users with authoritative information from established sources on how to participate in the election. This follows a more general strategy of promoting information from trusted sources as a form of “counter-speech” to undercut misinformation. For instance, Facebook has launched a “Voting Information Center” as a feature at the top of users’ feed to spotlight information about the election that is aimed at combating misinformation (e.g., promoting sources that explicitly state that mail-in voting is safe and voter fraud is exceedingly rare). Other platforms have taken similar approaches. TikTok has launched an in-app “Voter Guide” and YouTube has introduced “information panels” linked to searches of political candidates and searches on how to vote.
Finding rights-respecting approaches to difficult policy problems is particularly challenging when it requires the need to balance competing rights against one another. This is the case for perhaps all online content moderation questions, but is especially acute when it comes to how platforms craft policies in the context of elections. Stakes are especially high during the electoral period, but they are high in ways that are predictable, and therefore allow platforms to prepare for challenges that may arise.

Open exchange of political ideas is critical for citizens to make informed voting decisions, and digital platforms now constitute an essential medium for this exchange. It is therefore important for platforms to create space for the free expression of political ideas, and to avoid excessive moderation of political speech. That said, there are legitimate reasons to restrict speech when it infringes on other rights, and such tradeoffs often arise in digital contexts.

In the context of elections this is most clearly highlighted by the potential for online speech to interfere with people’s rights to freely and fairly elect their government. The risk here is that poorly enforced policies, or algorithmic recommendation systems that promote misinformation about elections or the electoral process, or which unevenly censor political speech, have the potential to threaten people’s right to participate in free and fair elections.

In the context of these kinds of tradeoffs, it’s important for platforms to be transparent about how they will respond when certain content threatens the health of the democratic process. This is something that many platforms are increasingly acknowledging—for instance, recent announcements from both Facebook and Twitter have included specific information about the conditions under which they will determine that election results are valid—but that is happening relatively late in the US election process, and has not been engaged as substantially in previous elections worldwide.

Additionally, an overall lack of transparency on how rules about electoral speech will be enforced, and a failure to craft clearly defined policies in advance, can lead to seemingly arbitrary decisions about electoral speech, and to uneven application of standards between groups. As platforms take steps to moderate increasingly ambiguous content (such as content which might indirectly impact electoral processes), the need for clear policies about the moderation process and how policies will be enforced is even more urgent.

The tradeoffs become even more complex in the context of speech from elected public figures and those running for office, particularly when these individuals say things which would otherwise not be allowed on the platform, like inciting violence or misleading users about voting. On the one hand, there is a public interest in voters knowing what their leaders or candidates say. On the other hand, misleading speech about voting procedures, safety or validity is likely to be even more harmful when it comes from candidates or elected officials. In these cases, platforms need to take measures to inhibit the ability of public figures to use these platforms to illegitimately undermine democratic processes, even if it means limiting some expression and the ability for people to access certain speech on the platform. It’s also critically important that enforcement on public figures is consistent. For example, if platforms claim that they will label misinformation about voting by public figures, but fail to do so each time a figure shares such misinformation, this could lend additional credence to inaccurate posts, because users expect that if the content was misleading it would be labeled.

One aspect of the right to free expression is the right to freely seek and receive information. It should therefore be easy for users to assess what kind of information they are unlikely to find or unable to share on platforms. To the extent that the platforms’ own proprietary algorithms determine what information people encounter about an election, platforms should consider giving users more insight into the impact of these algorithms on their informational experience, or allowing users to opt out of algorithms and view content, for example, chronologically (something Twitter already allows).
RECOMMENDATIONS FOR PLATFORMS IN FUTURE ELECTIONS

The 2020 US elections have motivated substantial changes in platform approaches to elections and election-related content, and platforms seem to be increasingly recognizing the role that they play in the electoral process. As platforms move forward and consider applying these new strategies to future elections, it is critical that they continue to evaluate the human rights impacts of platform rules during these periods. The following are recommendations for platforms in taking a rights-respecting approach to online content in future electoral periods:

1) Platforms should establish specific policies on how they will approach election-related content well in advance of the electoral period. While platforms will need to make real-time adjustments as specific unforeseen issues arise, foundational policies should be in place early on so that candidates and users have a clear understanding of the baseline rules that will govern their speech.

2) Platforms should consistently enforce these electoral content policies. While perfect enforcement is not possible, consistent enforcement should be a priority. Without consistent enforcement, the platforms risk creating the appearance of bias or favoritism and in turn, undermining confidence in their ability to fairly adjudicate and enforce their decisions during the electoral period.

3) Platforms should be transparent about their decision-making processes as they interpret and enforce their policies. This transparency into the process can give users a clearer understanding of how decisions are made.

4) Platforms should consider alternatives besides exemption from some or all moderation for public figures. For instance, legitimate public interest in preserving content from public figures (which would be removed under platform policies if posted by ordinary users) might be more appropriately accomplished by making such newsworthy removals accessible in a searchable database of such content available to journalists, researchers and the public.

5) If platforms allow political advertising, this advertising should be subject to restrictions on misinformation that are at least as stringent as those applied to other advertising or speech on the platform. If other advertising is prohibited from false claims, then political advertising should be subject to similar prohibitions. Advertising should not become a way to get around other moderation rules.

6) Platforms should use the range of tools at their disposal as a way to balance the protection of competing rights against each other. For example, taking algorithmic approaches to limit the spread of misinformation, or strategies to introduce friction into the sharing process, can allow platforms to both minimize the impact of harmful content while better protecting free expression.

DIVING DEEPER

Evaluating Platform Election-Related Speech Policies, Election Integrity Partnership
Freedom of Expression and Elections in the Digital Age, OHCHR
Freedom and Accountability A Transatlantic Framework for Moderating Speech Online, Transatlantic Working Group
Election 2020: Content Moderation and Accountability, Stanford Institute for Human-Centered Artificial Intelligence
Facebook Has Made Lots of New Rules This Year. It Doesn’t Always Enforce Them, Wall Street Journal

AUTHORS

Megan Metzger is a Research Scholar and Associate Director of Research for the Global Digital Policy Incubator. She completed a PhD in Politics at NYU as a member of the Social Media and Political Participation Lab and her research focuses on how technology impacts rights and political behavior.

Rosanna Kim is a Research Assistant at the Stanford Global Digital Policy Incubator. She focuses on the human rights implications of content moderation by internet platforms and governments. She completed her MA in international policy at Stanford and holds a BA in political science and history from Swarthmore College.

*With research assistance from Barby Hernandez and Sarah Lee.

This snapshot was created in partnership with the International Center for Not-for-Profit Law.