YouTube Updates its Policy on Election Misinformation

Richard Mackenzie-Gray Scott   2023-06-08T07:29:27

Last Friday, YouTube announced that it 'will stop removing content that advances false claims that widespread fraud, errors, or glitches occurred in the 2020 and other past US Presidential elections'. This development has upsides and downsides, a few of which are worth sketching out, and all of which further accentuate why the US constitutional framework regarding online platform regulation requires updating. The nature of this update requires transcending a governance approach of overreliance on expecting good faith self-regulation by companies providing these intermediaries.

Downsides

There are many reasons why leaving misinformation circulating online is a problem. These mainly concern what the spread of misinformation can contribute to down the line, whether vaccine hesitancy (and related death) or electoral distrust (and related destruction). Whatever these downstream effects may be, they can share two common denominators that help highlight why leaving misinformation circulating online is a potentially bad decision. The first consists of misunderstanding the marketplace of ideas metaphor and not accounting for its limitations in online information environments. The second concerns underestimating information operations targeting elections and their relationship to the spread of misinformation.

Online marketplaces of attention where ideas are unevenly distributed and weighted

The marketplace of ideas metaphor underpinning the US constitutional framework attempting to govern expression and thought does not account for how ideas are distributed and weighted across online networks. Ideas can only compete if they are known and rationally weighed against prior knowledge. Algorithmic curation by online platforms and the cognitive quirks of users combine with the result that people are not aware of different sources of information, and if made aware, process and respond to them differently. There are measures offering a corrective to this dynamic, such as those providing for 'informational osmosis', which is when alternative sources of information are introduced into the informational bubble of a user. However, the underlying assumption that truth will ultimately emanate from falsehood if ideas are left to compete in online information environments is faulty, especially considering platforms decide *what* ideas receive *whose* attention. A recent study found YouTube's recommendation algorithm presented content featuring narratives questioning the previous US Presidential election to users that were more likely sceptical of its legitimacy. If algorithmic curation continues operating in such a way, then leaving electoral misinformation online risks bolstering belief in it, including from people that may not be likely, at least initially, to lend it credence. Increases in exposure to such misinformation is made more problematic if no alternative

explanations are offered that might assist in users switching out the false information in their belief systems. Democratic debate rests in part on people having access to the same sources of information, not differentiating exposure to those sources depending on their data profile.

Information operations that saturate online spaces with misinformation

Further complicating choices to leave election misinformation online are actors looking to influence user behaviour through its spread. An old problem with new components, information operations occur today in attempts to forge the future according to their planners' preferences. For example, a company or government may deploy bots across platforms before an election for the purpose of influencing the result and/or behaviour towards it. This 'computational propaganda' can result in spikes of misinformation on those platforms. The false information that initially begins as disinformation shared by a small number of non-human accounts can turn into misinformation being shared by a large number of users. This situation can be compounded by 'reverse censorship', part of which consists of users coordinating to increase the amount and reach of content underpinned by a particular partisan stance. Should this content contain falsehoods, there is the risk that concentrations of accurate information become diluted on hosting platforms. A Facebook report claims that in information operations targeting elections, 'authentic voices typically outweigh inauthentic attempts to manipulate public debate'. But by not removing electoral misinformation, YouTube is banking on any increase of it not contributing to a decrease in the amount of accurate information to which users are exposed. While the probabilities of how this gamble unfolds are unknown and perhaps immeasurable, the inescapable aspect of it is the vulnerability of the process to manipulation, where parties interested in particular electoral outcomes can assist misinformation attaining greater reach and impact than accurate information.

Upsides

Despite these (mitigable) downsides, the latest YouTube policy is not necessarily bad when considering the interconnected human rights of expression and thought. Further appreciation of this perspective can be gained when reflecting on their link to perhaps the most significant factor in countering online misinformation: trust.

Divergent thought and expression are crucial to democracy founded on pluralism

Content removal on the grounds that it is misinformation can be arbitrary and thus 'suspect under a liberal commitment to free expression'. Aggravating factors here are when these decisions contain no clarity about what constitutes misinformation, how such determinations are made, and any inconsistent enforcement. Measures short of removing misinformation from online platforms are preferable from this perspective because they grant comparatively more agency to users. By allowing people on YouTube the opportunity to seek out, receive, and impart what may be electoral misinformation, the platform is attempting to respect freedom of expression and freedom of thought. This motivation is implied in the marketing of their change in policy: 'The ability to openly debate political ideas, even those that are controversial

or based on disproven assumptions, is core to a functioning democratic society–especially in the midst of election season'. Whether and how less limitations being placed on the exercise of expression and thought necessarily corresponds to them being respected is another, meatier question. Yet more significant than arguments revolving around the extent to which particular content moderation decisions regarding misinformation do or do not respect these two human rights, is user trust in these processes.

Dispelling misinformation depends on trust

[Research from The Royal Society](#) helps show why content removal is a poor option to effectively managing misinformation, notably because it may 'exacerbate feelings of distrust and be exploited by others to promote misinformation content'. Misinformation is, by definition, created and spread by people that have a knowledge deficit about the content at issue and lack malicious intent when sharing it. Feelings of innocence and offence in connection to accusations of promulgating misinformation, which are implicit in content removal decisions, may lead to users refusing to consider, never mind believe, sources of information that contradict misinformation. Trust in sources of information and in the systems that disseminate them depend on *how* communication occurs perhaps as much [if not more](#) than *what* is being communicated – and takedowns are a form of the how. If the connection between media and political trust is [being driven by](#) public sentiment against elites, then those who are considered to comprise these groups need to exercise care in how they express their views so as not to come across as infantilising, patronising, or the like. Very large online platforms and their overlords arguably fall into the category of an elite group, meaning their removal decisions can be perceived disapprovingly by users, the negative feelings connected to which may be heightened if users disagree that the content was misinformation.

Related to these factors is the Streisand effect, which describes attempts to restrict access to information that have the paradoxical effect of leading to it receiving considerably more attention than it otherwise would have if left be. Censorship can be counterproductive. First, because the associated user distrust may mean more people believing conspiracy theories surrounding the applicable content. Second, because when accounting for the switching costs to users of online platforms, removing content may have no short- or long-term impact on reducing the spread of misinformation, and could even result in its increase across other platforms. The decision of users to change platforms is made easier when content they are wanting to access is available on one intermediary but not another. This may well be another reason behind YouTube reversing its previous policy on US Presidential election misinformation, as resorting to takedowns risks the platform losing users, and thus the related revenue that their engagement generates.

A constitutional update is in order

The U-turn from YouTube regarding its policy on US Presidential election misinformation is yet another instance marking just how much societies are at the mercy of powerful online platforms. The unpredictability of their decisions, reliance

on unaccountable algorithmic systems of information dissemination, and wide scope for abuse of their role, render the notion of democratic oversight illusory. These issues stem from the considerable extent to which platform conduct is discretionary under the US constitution. Whether originating from new legislation or amendments to existing legislation, an update is in order. Even if US lawmakers struggle to understand the ins and outs of platform regulation, there is appetite for reform, including from Democrats and Republicans alike. The status quo is not only putting courts in difficult positions regarding content moderation, but also failing to realise the value of governing through a participatory system providing checks and balances on power. Regulations in the US securing wealth extraction have allowed online platforms such as YouTube to grow into the positions of hegemony they now hold. It is past time to rebalance the scales. Let there be less 'whack-a-mole' approaches to content moderation, spurred by legitimising (possibly) deceptive and performative mechanisms of accountability that rely on appearing juridical. The procedural aspects of platform regulation require more attention. And as to the substance of the US free speech culture emboldened by its constitution, history and recent research leave food for thought.

After the US Bill of Rights entered into force, the Alien and Sedition Acts of 1798 criminalised speech containing 'false, scandalous and malicious' content about public institutions and officials, punishable 'by a fine not exceeding two thousand dollars, and by imprisonment not exceeding two years'. The First Amendment has not always been such a prominent consideration in the governance of public life. If US citizens now prefer 'quashing harmful misinformation over protecting free speech', then it can be questioned whether this constitutional provision and accompanying jurisprudence should continue to retain as much influence in debates and decisions about content moderation.

The people working at YouTube may genuinely care about electoral integrity. Even so, the preferences of its owning company (Alphabet) are guided and shaped by the dictates and incentives of a market system favouring profit maximization. These preferences are not the same as those of the public. With evidence showing further incongruence between these preferences, should those of platforms be taking precedence? While YouTube and its compatriots may again update their election misinformation policies as the next US Presidential election looms, that US society and beyond are all left waiting to see what happens is unsatisfactory.