Coronavirus and the frailness of platform governance

Magalhães, João Carlos; Katzenbach, Christian

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Coronavirus and the frailness of platform governance

João Carlos Magalhães, Alexander von Humboldt Institute for Internet and Society, Berlin, Germany, joao.magalhaes@hiig.de
Christian Katzenbach, Alexander von Humboldt Institute for Internet and Society, Berlin, Germany, christian.katzenbach@hiig.de

PUBLISHED AT Internet Policy Review, *29 Mar 2020, https://policyreview.info/articles/news/coronavirus-and-frailness-platform-governance/1458*

Major health crises, historian David S. Jones recently reminded us "put pressure on the societies they strike". And this strain, he points out, "makes visible latent structures that might not otherwise be evident". Something similar is happening now. As the novel coronavirus pandemic quickly morphs into an unprecedented global calamity, issues that not long ago seemed acceptable, fashionable, and even inescapable - such as fiscal austerity and science-scepticism, are increasingly called into question. Unsurprisingly in an era dominated in many ways by 'Big Tech', the pandemic has also helped to foreground how contestable – and, we argue, utterly frail – *platform governance* is. By this expression we mean the regimes of rules, patterned practices and algorithmic systems whereby companies govern who can see what in their digital platforms.

While all eyes are on public health, the larger economic wellbeing and other emergencies, platform governance is far from being superfluous. In a moment where we all heavily depend on digital services to receive and impart news to make sense of the current situation, the way companies such as Facebook and YouTube manage the content on their platforms play an obvious role in how the very pandemic evolves. More than influencing the crisis, though, these services have already been changed by it.

Sending moderators home: a sharp turn to AI in content moderation

Consider two recent developments.

As the outbreak escalated, Facebook and YouTube announced last week that decisions on whether to keep or take down certain posts would rely less on human moderators (who would be sent home to avoid contamination) and more on algorithmic systems. Increased automation, they admitted, would lead to more "mistakes" in the management of content in the massive public spaces they privately control. Google (who owns YouTube) said on March 16 that "there may be an increase in content classified for removal during this time". Facebook sounded a little more defensive and vague, when arguing on March 16 that "we may see some longer response times and make more mistakes as a result" but that this shouldn't "impact people using our platform in any noticeable way".

Another move was made by Twitter. Responding to growing concerns over misleading content about the pandemic, the platform announced in a corporate post on March 16 that it would adopt a draconian moderation policy in regards to coronavirus-related posts. From then on, Twitter would request the removal of all "content that increases the chance that someone contracts or transmits the virus. This apparently includes even tweets suggesting that "social distancing is not effective".

Even when taken at their face value, these changes should raise an eyebrow. While it is commendable to acknowledge that automated content moderation might produce more "mistakes", Google and Facebook's announcements fall short of explaining the various problems involved in the use of algorithmic systems to perform a task that reasonable humans still mightily struggle to agree upon. To begin with, it is unclear what exact "mistakes" this automation will produce. Facebook users quickly denounced that posts with legit information about the pandemic were taken down as spam -- what the company called a mere "bug".

As one of us argued in a recent co-authored paper in *Big Data & Society*, an almost fully automated system of content moderation bears the dangers of hiding the political nature of decisions over content. What if these moderation systems achieve their overarching aim by becoming an infrastructure that smoothly operates in the background, that is taken for granted? Such infrastructures of public speech obscure their inner workings and the fundamentally political

nature of speech rules being executed by potentially unjust software at scale.

The politics of decisions over content in a pandemic crisis

Twitter's decision on content related to the novel coronavirus, for example, seems to assume a level of conceptual clarity and institutional legitimacy that simply do not exist. Making sense of evolving pandemics like this one is an extraordinarily complex task, even for epidemiologists. For instance: some weeks ago, many experts were telling us that social distancing should mainly apply to sick individuals, only to realise (after some research) that asymptomatic people could also transmit the virus. If experts are unsure on what to do, why should we trust Twitter with the one-sided ability to say which content can fuel the transmission of the virus?

Less than 24-hours after the new policy was announced, the platform gave us good reasons to be concerned. Elon Musk, the powerful CEO of Tesla, who has repeatedly downplayed the seriousness of the pandemic, tweeted the false information that "kids are essentially immune" to the new coronavirus. This might appear a blatant example of what the platform had just forbidden. But the post was not removed. "It does not break our rules", Twitter declared after reviewing the "overall context and conclusion of the Tweet".

Origins of frailness: concentrated production chains, unstable rules, unaccountable decisions

It is not the first time of course that Twitter appears to protect a powerful billionaire, as its seeming complacency with Donald J. Trump's behaviour suggests. Indeed, the particular issues that the current coronavirus crisis seem to underscore point to a much more fundamental problem: companies' content governance regimes depend on remarkably frail arrangements.

This frailness is in part related to how concentrated content moderation "production chains" are. The current turn to automation, for instance, is caused by the fact that many human moderators are not allowed to work from home. This might seem surprising. Aren't technology companies able to design safe systems for this kind of job to be done remotely? As explained by Sarah T. Roberts, an UCLA (University of California, Los Angeles) assistant professor, remote content moderation might be precluded by "constraints like privacy agreements and data protection policies in various jurisdictions". A disproportionate amount of the distressful labour that goes into moderation is exerted by multitudes of low-paid individuals in poor countries. In fact, the current shortage of moderators appears to be directly linked to the quarantine of a particular group of workers in Manilla, she says. "What is supposed to be a resilient just-in-time chain of goods and services… may, in fact, be a much more fragile ecosystem in which some aspects of manufacture, parts provision, and/or labor are reliant upon a single supplier, factory, or location."

Another facet of platform governance's frailness regards the instability of companies' internal rules. Sudden and reactive policy changes, like Twitter's new coronavirus policy, are a constant. "When you look at a site's published content policies", says a representative from a platform quoted in a book by Cornell University's Tarleton Gillespie, "there's a good chance that each of them represents some situation that arose, was not covered by existing policy, turned into a controversy, and resulted in a new policy afterward".

Recently, we at the HIIG examined how 'Twitter Rules' (the platform's community guideline) changed since 2009. Our analysis found over 300 changes in directives, terminology and classification of regulations. Many of these shifts were obviously associated with specific external events, such as the 2016 US presidential election and the ongoing ethnic conflict in India. Others appeared to reveal the seemingly erratic ebbs and flows of a company unsure of how to exert its enormous powers, e.g. the incremental complexification and then sudden simplification of "spam" definitions. Overall, these changes seem to document Twitter's slow and reluctant emergence as an explicitly political institution.

Finally, the suspicions triggered by the way in which Twitter apparently overruled its own policy not to punish Elon Musk evokes platform governance's perennial political fragility. That is, the lack of stable transparency channels whereby the rest of society can minimally understand companies' policymaking and technology design and management. The decision-making of major social media platforms remains essentially unaccountable, often the prerogative of a clique of executives and employees whose concerns, methods and (likely) disputes have been essentially hidden from minimal public scrutiny1. While fiercely defended by companies as key to their business model,

this transparency deficit arguably weakens their legitimacy, increases external criticism and eventually leads these companies to experiment with new governing practices. Facebook, for instance, now seems to be implementing its own "Supreme Court". Whether this initiative will flourish, and for how long, is unclear.

Platform governance after the novel coronavirus

Will such frailness resist? Can we expect platform governance to emerge from this pandemic as more reliable, stable and democratic?

The frailness we described so far maintained a complex relationship to previous crises. Much of platform governance regimes originated as adaptive reforms, hasty solutions to placate external criticism and instabilities. Take the unstable internal policies and the escalation of content moderation with cheap human labour – largely done after the so-called "techlash". On the other hand, unaccountable decision-making has continually hindered our ability to understand the extent to which companies could be indeed involved in recent watershed events. The use of platforms by Russia's disinformation agency during the 2016 US presidential election, for instance, was unveiled by journalists, academics and judicial investigations. Companies like Facebook initially denied and deflected any criticisms.

The last years taught us that platforms are unlikely to truly enhance, on their own, governance regimes that, while frail, are also profitable. They will have to be pressured. And this pressure will only be strong enough to promote any structural change if platforms are shown to have played a part in the pandemic. What was the role that disinformation circulating online played in the mushrooming of the cases? Did companies abate or worsen the problem? Should they be indirectly involved in the death of dozens of thousands of people? It is likely that the magnitude of the trouble will finally prove too high for companies to weather. It remains to be seen how the opacity of an increasingly automated content moderation system may affect this assessment.

However, if this crisis ends up being a moment of further consolidation of Big Tech's social power, as some predict, their governance arrangements will probably go unchallenged for a long time. Or, perhaps worse, companies might use this crisis to normalise money-saving solutions that in normal times would be ethically unacceptable – think of the "mistakes" generated by the further turn to AI, peddled as the minor cost of grim trade-offs.

To say that shocks often work as catalysts of structural changes does not tell us the direction of the transformation. There is no guarantee that any lasting change will be in the public interest. Policymakers, journalists and researchers must redouble their accountability efforts. The governance regimes being renegotiated now are poised to be an even more central structure in the world that will emerge from this cataclysm.

Footnotes

1. See however the recent pioneering study (PDF) of our colleagues Matthias C. Kettemann and Wolfgang Schulz on Facebook's private policy making.