

The road to hell is paved with good algorithms: The case for deactivating recommendation algorithms in the political sphere

By Jonas Kaiser

Algorithms, and recommendation algorithms in particular, are deeply ingrained in our [networked public sphere](#). Facebook recommends us new friends, pages to like or groups to join, Google websites that fit our search interests, Twitter people to follow or topics to check out, Amazon products to buy, Spotify music to listen to, and YouTube videos to watch.

According to Alexa, the most visited websites in the U.S. are Google, YouTube, Amazon, Yahoo, and Facebook. Recommendation algorithms are an integral part of each of those websites. As [Eli Pariser](#), [Safiya Umoja Noble](#), [Frank Pasquale](#), and many others have argued, there are myriad reasons to be concerned about algorithms' integral role in our daily lives. These include the creation of homogenous communities without our knowledge; reproduction of racism and misogyny; and the concealment of algorithmic decisions to begin with. Against this background, as well as my own research on the far-right, disinformation, and algorithms, I argue that social media companies need to deactivate their recommendation algorithms in the political sphere. I structure this demand around our knowledge of how machine learning works, the pitfalls of automated content curation, how corporate goals can run counter to the public good, as well as findings from my own research.

When thinking about recommendation algorithms, we need to think about machine learning, and statistical probability models. After all, this is what algorithms are. Yet, as British statistician [George Box famously said](#): "All models are wrong, but some are useful." As [Momin Malik](#) highlights convincingly, these models are approximations, and there are numerous ways in which they can fail or have shortcomings. On a more applied note, [Harini Suresh and John Guttag](#) identified five biases that compromise algorithms: historical bias, representation bias, measurement bias, aggregation bias, and evaluation bias. These biases can lead to problematic outcomes that might reproduce issues such as racism, misogyny, and more. In short: algorithms, no matter how good, will always have limitations.

This problem is only exacerbated when focused on the supply side — i.e., the content on a platform. As my colleague [Adrian Rauchfleisch and I have highlighted](#), far-right content that heavily features racism or disinformation plays a significant role in the German as well as the American YouTube sphere. When comparing the prominence of far-right actors on YouTube with [general findings for the U.S. networked public sphere](#), the extreme political fringe seems overrepresented on YouTube. Similarly, in Germany, the [YouTube political community](#) mostly consists of far-right and conspiratorial channels. This, then, highlights that political communities on YouTube are hardly representative of the general media discourse and seem to favor more radical voices.

But as [Michael Golebiewski and Danah Boyd](#) highlight, even if that were not the case, algorithms face an inherent issue: data voids. Data voids are, in short, gaps in the content that a platform can recommend. For example, this can occur when a specific search term suddenly gains popularity. These data voids, however, can be abused by malicious actors who want to spread disinformation; recommendation algorithms cannot *not* recommend content. Indeed, they are limited to content on their platforms. This inherent need to recommend thus can feature harmful content; this is especially so when content on the platform is already harmful. As my co-authors and I show in a forthcoming study on Zika in Brazil, even when YouTube curated the search results for videos on the Zika virus, misinformation was still present throughout the results and recommendations. This, then, indicates that even when platforms attempt to curate their recommendations, algorithms will nevertheless uncover and recommend harmful content.

Add to this that algorithms that are usually the property of companies and thus, as Pasquale, highlights “black boxes.” This means that we can inherently only see, measure, or interact with an algorithm’s output and can only guess on how the algorithm ended up with its final recommendations. While every now and then we get an idea of some of the factors that contribute to a [platform’s algorithms](#), the general audience, as well as the platform’s content creators, are left in the dark. On some platforms like YouTube (but also recently TikTok), the algorithm has thus even a “celestial” quality, as [content creators’ success](#) is dependent on it.

I argue, however, that we don’t need to know what goes into the algorithm to understand that their objectives are at odds with the public good and a [utopian version of the public sphere](#). Indeed, from everything that we know, [algorithms are optimized on user behavior](#) and especially on how much time is spent on the platform. And while companies profit off of users’ prolonged stays on their platforms, it is unlikely that users are profiting to the same extent.

Indeed, what keeps people engaged, i.e., viewing, commenting, etc., can to some extent be traced back to negative content. As we know from [studies on user comments](#), people tend to write user comments that are more uncertain, negative, and controversial. In addition, there is a reason why [YouTube’s study](#) focuses on users’ satisfaction; presumably because engagement on YouTube is not driven by content that is agreeable but rather controversial. In other words: What is good for the company is not necessarily in service of the public good.

Which brings me to my final point: the effect of algorithms. Indeed, little is known so far in terms of the effect of recommendation algorithms. Yet, [a study that I conducted](#) looked at the user comments in over 100 German far-right channels and examined whether we could identify activity patterns over time. Indeed, we were able to show that the community grew more central over time, indicating that the users that, at first, only commented under one channel eventually also commented under videos from

other related channels. And while we don't know whether this finding can be explained with YouTube's algorithms alone, it is important to note that YouTube claims its algorithms drive [70% of the traffic on the site](#). In this context, [we have argued](#) that YouTube's recommendation algorithms can cause a digital Thomas theorem that normalizes radical content and can, effectively, nudge people towards more problematic and disinforming content.

Recommendation algorithms are, in my opinion, unfixable. There is no doubt that they, in general, work and can be quite useful. Indeed, most content on Facebook or YouTube is not political and, in these contexts, one might have fewer issues with recommendation algorithms. Yet, even in these supposedly benign contexts, algorithms can cause harm. [While conducting research on YouTube in Brazil](#), my co-authors and I stumbled on what *The New York Times* eventually called "[On YouTube's Digital Playground, an Open Gate for Pedophiles](#)."

While recommendation algorithms are supposedly neutral, neither the people creating them nor the people using or training them are. For YouTube's recommendation algorithm, people watching content is people watching content and the algorithm attempts to optimize on their viewing behavior; in this case, videos of children. And while this might be an extreme example of a recommendation algorithm causing real harm, [YouTube acted swiftly and deactivated its recommendation algorithm](#) when videos included children.

In this piece I am arguing for a similar step for political content. As I have shown above, no matter how much work platforms pour into their algorithms, they will always have limitations and will always need curation. Combining these imperfect algorithms with a highly skewed group of content creators, then, must end badly. No matter how much you tweak and optimize the algorithms, if content is problematic, the algorithms will recommend it. Deliberate attempts at "gaming" the algorithm and pushing disinformation, coupled with humans being drawn to controversial and negative content, means that as long as recommendation algorithms exist, problematic content will surface and be recommended. This is especially so if algorithms are designed to keep people on the platform, even when it is not for their own good. Finally, these algorithms can contribute to the normalization of extreme and disinforming content and nudge people to more radical communities.

I am not saying that these platforms should only remove specific recommendations. No. I argue that we need to get rid of all recommendations for political content. This is not about whether content can or cannot exist on a platform. This is about whether recommendation algorithms can be "saved." I argue that they can't and only demand platforms to take the same steps that YouTube did when made aware of their algorithms pushing videos of children to pedophiles: Deactivate the recommendation algorithms.

Jonas Kaiser is an assistant professor of communication, journalism, and media at Suffolk University.