

TO: Kenan Institute for Ethics & the Duke Community at Large

FROM: Jessica Edelson, '22 and Niharika Vattikonda, '22

DATE: April 2020

SUBJECT: TikTok's Physiognomic Bubbles and Algorithmic Bias Amplify Messages of Hate

With over 1 billion users, TikTok has gained notoriety for its endless feed of bite-sized content. The average user spends at least an hour each day watching and creating short clips, ranging from choreographed dance routines to political rants. Whether an individual has a formal account on the platform or not, TikTok's algorithm generates content recommendations, churning out videos in a manner that feels eerily robotic and enticingly personal. The algorithm measures the length of time a user watches videos and monitors how that user engages with content through the like, comment, and share features. The result is a recommendation algorithm that is highly predictive but also rashly biased.

The TikTok algorithm is representative of a key emerging technology challenge: algorithmic discrimination. According to an [experiment conducted by UC Berkeley researcher Marc Faddoul](#), TikTok creates physiognomic bubbles that adapt recommendations not only based on gender and ethnicity, but also "hair style, body profile, age, how un(dressed) that person is, and even whether they have visible disabilities." This bias is the result of unrepresentative data training sets often out of the programmer's control and lackluster [moderation policies](#).

The consequences of a feedback loop that places users within echo chambers cannot be ignored. TikTok's moderators are unable to screen every single video on the platform, and this has previously led to videos of [white supremacy and Nazism on the platform](#) and [caste-based violence in India](#). In the case of white supremacist content, the current TikTok model would push the harmful messages out to white-appearing viewers. These viewers could be particularly inveigled by the hate-speech due to the physiognomic bubbles: their own identity would be mirrored in that of the content-deliverer. Additionally, other users would not be shown the content and would therefore not have the opportunity to flag it for moderator review.

TikTok may not be able to save us from what Syracuse professor Whitney Phillips [refers to](#) as "algorithmic content hell" in which "people naturally-self select...within entertainment ecosystems." But the company should aim to prevent the spread of hateful content, particularly given that their user base consists of younger individuals who may be more susceptible to the nuanced targeting carried out by extremists online. Though TikTok is not the direct creator of harmful content, operating the recommendation algorithm as a "black box" and allowing for the incitement of violence and hatred is unethical. TikTok has a moral obligation to inform users of how their feeds are being manipulated.

We recommend that TikTok move the "Not interested" button so that users are aware of this option and able to exercise it appropriately. Currently, that button is hidden within the share

feature, which feels counterintuitive and is tucked out of sight, a manipulative move by the creators of the app's user interface. TikTok should move this button next to the standard like, comment, and share buttons so that users have an equal choice in demonstrating their like or dislike of content. The TikTok interface could also be adjusted to allow users to swipe left or right, akin to the dating app Tinder, in order to indicate their content preferences. This simple modification would be a practical way to start providing users with increased autonomy and a more democratic TikTok experience.

TikTok should also provide algorithmic transparency so users can understand how their interactions with the app drive the recommendation algorithm—this would involve TikTok posting the algorithm's source code. However, the source code itself may not contain the clues related to algorithmic discrimination—those clues will be found in the training data. TikTok should also provide an annual report on the breakdown of the training data set: this report would include a breakdown of how various populations are represented in the training data, without requiring the company to open up their entire data set, which would be both taxing and impractical. Explainable AI (XAI) can be a valuable tool to help users break down the meaning of this data—[XAI allows users to understand the rationale behind the decisions made by AI](#), preventing the formation of “black box” technology that confuses users.

Recommendation algorithms are a fundamental part of most social networks and search algorithms, but in TikTok's case, this recommendation algorithm only perpetuates the siloing of users within the platform. This algorithm allows content featuring white supremacy, Nazism and caste-based discrimination, among others, to propagate through the platform, allowing nefarious actors to target viewers by appearance and by race. It is unethical for TikTok to continue to operate the recommendation algorithm as a black box, knowing that operatives are taking advantage of TikTok's younger user base. In a digiverse of mindless scrolling, product design changes would allow users to have more autonomy, and algorithmic transparency coupled with explainable AI would result in a more democratic experience for users.