

How TikTok Reads Your Mind: the media equation

Smith, Ben. *New York Times*
(Online) New York Times Company. Dec 5, 2021.

The New York Times

Full text Details 3 Cited on ProQuest

Full text

 Listen

It's the most successful video app in the world. Our columnist has obtained an internal company document that offers a new level of detail about how the algorithm works.

There are four main goals for TikTok's algorithm: 用户价值, 用户价值 (长期), 作者价值, and 平台价值, which the company translates as "user value," "long-term user value," "creator value," and "platform value."

That set of goals is drawn from a frank and revealing document for company employees that offers new details of how the most successful video app in the world has built such an entertaining — some would say addictive — product.

The document, headed "TikTok Algo 101," was produced by TikTok's engineering team in Beijing. A company spokeswoman, Hilary McQuaide, confirmed its authenticity, and said it was written to explain to nontechnical employees how the algorithm works. The document offers a new level

of detail about the dominant video app, providing a revealing glimpse both of the app's mathematical core and insight into the company's understanding of human nature — our tendencies toward boredom, our sensitivity to cultural cues — that help explain why it's so hard to put down. The document also lifts the curtain on the company's seamless connection to its Chinese parent company, ByteDance, at a time when the U.S. Department of Commerce is preparing a report on whether TikTok poses a security risk to the United States.

If you're among the billion people (literally!) who spend time on TikTok every month, you're familiar with the app as 2021's central vehicle for youth culture and online culture generally. It displays an endless stream of videos and, unlike the social media apps it is increasingly displacing, serves more as entertainment than as a connection to friends.

It succeeded where other short videos apps failed in part because it makes creation so easy, giving users background music to dance to or memes to enact, rather than forcing them to fill dead air. And for many users, who consume without creating, the app is shockingly good at reading your preferences and steering you to one of its many "sides," whether you're interested in socialism or Excel tips or sex, conservative politics or a specific celebrity. It's astonishingly good at revealing people's desires even to themselves — "The TikTok Algorithm Knew My Sexuality Better Than I Did," reads one in a series of headlines about people marveling at the app's X-ray of their inner lives.

TikTok has publicly shared the broad outlines of its recommendation system, saying it takes into account factors including likes and comments as well as video information like captions, sounds and hashtags. Outside analysts have also sought to crack its code. A recent Wall Street Journal report demonstrated how TikTok relies heavily on how much time you spend watching each video to steer you toward more videos that will keep you scrolling, and that process can sometimes lead young viewers down dangerous rabbit holes, in particular toward content that promotes suicide or self-harm — problems that TikTok says it's working to stop by aggressively deleting content that violates its terms of service.

The new document was shared with The New York Times by a person who was authorized to read it, but not to share it, and who provided it on the condition of anonymity. The person was disturbed by the app's push toward "sad" content that could induce self-harm.

The document explains frankly that in the pursuit of the company's "ultimate goal" of adding daily active users, it has chosen to optimize for two closely related metrics in the stream of videos it serves: "retention" — that is, whether a user comes back — and "time spent." The app wants to keep you there as long as possible. The experience is sometimes described as an addiction, though it also recalls a frequent criticism of pop culture. The playwright David Mamet, writing scornfully in 1998 about "pseudoart," observed that "people are drawn to summer movies because they are *not* satisfying, and so they offer opportunities to repeat the compulsion."

To analysts who believe algorithmic recommendations pose a social threat, the TikTok document confirms their suspicions.

“This system means that watch time is key. The algorithm tries to get people addicted rather than giving them what they really want,” said Guillaume Chaslot, the founder of Algo Transparency, a group based in Paris that has studied YouTube’s recommendation system and takes a dark view of the effect of the product on children, in particular. Mr. Chaslot reviewed the TikTok document at my request.

“I think it’s a crazy idea to let TikTok’s algorithm steer the life of our kids,” he said. “Each video a kid watches, TikTok gains a piece of information on him. In a few hours, the algorithm can detect his musical tastes, his physical attraction, if he’s depressed, if he might be into drugs, and many other sensitive information. There’s a high risk that some of this information will be used against him. It could potentially be used to micro-target him or make him more addicted to the platform.”

The document says watch time isn’t the only factor TikTok considers. The document offers a rough equation for how videos are scored, in which a prediction driven by machine learning and actual user behavior are summed up for each of three bits of data: likes, comments and playtime, as well as an indication that the video has been played:

$$P_{like} \times V_{like} + P_{comment} \times V_{comment} + P_{playtime} \times V_{playtime} + P_{play} \times V_{play}$$

“The recommender system gives scores to all the videos based on this equation, and returns to users videos with the highest scores,” the document

says. "For brevity, the equation shown in this doc is highly simplified. The actual equation in use is much more complicated, but the logic behind is the same."

The document illustrates in detail how the company tweaks its system to identify and suppress "like bait" — videos designed to game the algorithm by explicitly asking people to like them — and how the company thinks through more nuanced questions.

"Some authors might have some cultural references in their videos and users can only better understand those references by watching more of the author's videos. Therefore, the total value that a user watches all those videos is higher than the values of watching each single video added up," the document says. "Another example: if a user likes a certain kind of video, but the app continues to push the same kind to him, he would quickly get bored and close the app. In this case, the total value created by the user watching the same kind of videos is lower than that of watching each single video, because repetitiveness leads to boredom."

"There are two solutions to this issue," the document goes on. "Make some assumptions, and break down the value into the value equation. For instance, in terms of repeated exposure, we could add a value 'same_author_seen,' and for the boredom issue, we could also add a negative value 'same_tag_today.' Other solutions besides value equation may also work, such as forced recommendation in users' for u feed and dispersion etc. For example, the boredom issue can be solved through dispersion."

Another chart in the document indicates that “creator monetization” is one of the company’s goals, a suggestion that TikTok may favor videos in part if they are lucrative, not just entertaining.

Julian McAuley, a professor of computer science at the University of California San Diego, who also reviewed the document, said in an email that the paper was short on detail about how exactly TikTok does its predictions, but that the description of its recommendation engine is “totally reasonable, but traditional stuff.” The company’s edge, he said, comes from combining machine learning with “fantastic volumes of data, highly engaged users, and a setting where users are amenable to consuming algorithmically recommended content (think how few other settings have all of these characteristics!). Not some algorithmic magic.”

Mr. McAuley added that he was a bit perplexed about why people were always asking him about TikTok.

“There seems to be some perception (by the media? or the public?) that they’ve cracked some magic code for recommendation, but most of what I’ve seen seems pretty normal,” he wrote.

And indeed, the document does much to demystify the sort of recommendation system that tech companies often present as impossibly hard for critics and regulators to grasp, but that typically focus on features that any ordinary user can understand. The Journal’s coverage of leaked Facebook documents, for instance, illustrated how Facebook’s decision to give more weight to comments helped divisive content spread. While the models may be complex, there’s nothing

inherently sinister or incomprehensible about the TikTok recommendation algorithm outlined in the document.

But the document also makes clear that TikTok has done nothing to sever its ties with its Chinese parent, ByteDance, whose ownership became a spasmodic focus at the end of President Donald J. Trump's administration in 2020, when he attempted to force the sale of TikTok to an American company allied with his administration, Oracle.

The TikTok document refers questions to an engineering manager whose LinkedIn biography says he works on both TikTok and ByteDance's similar Chinese app, Douyin, offering a glimpse at the remaining global element of an increasingly divided tech industry, the engineering talent. According to LinkedIn, the engineering manager attended Peking University, received a master's degree in computer science at Columbia University and worked for Facebook for two years before coming to ByteDance in Beijing in 2017. The document is written in clear, but nonnative, English, and comes from the perspective of the Chinese tech industry. It makes no references, for instance, to rival American companies like Facebook and Google, but includes a discussion of "if Toutiao/Kuaishou/Weibo have done something similar, can we launch the same strategy as they have done?"

TikTok's development process, the document says, is closely intertwined with the process of Douyin's. The document at one point refers TikTok employees to the "Launch Process for Douyin

Recommendation Strategy,” and links to an internal company document that it says is the “same document for TikTok and Douyin.”

TikTok employees are also deeply interwoven into ByteDance’s ecosystem. They use a ByteDance product called Lark, a corporate internal communications system like Slack but with aggressive performance-management features aimed at forcing employees to use the system more. There is, for instance, a graphic that tells you whether you have performed actions — like opening messages — more or less than your co-workers, according to screenshots I was given.

Concern about Chinese consumer technology is bipartisan in the United States. President Trump’s executive order attempting to ban the app in August 2020 warned that TikTok’s “data collection threatens to allow the Chinese Communist Party access to Americans’ personal and proprietary information.” The Chinese government could “build dossiers of personal information for blackmail, and conduct corporate espionage,” it said. That ban stalled in court and faded after the presidential election. President Biden rescinded the executive order, but his administration then announced its own investigation into security threats posed by TikTok, with an unnamed senior administration official telling reporters that China was “working to leverage digital technologies and American data in ways that present unacceptable national security risks.”

In an emailed statement, Ms. McQuaide said that “while there’s some commonality in the code, the TikTok and Douyin apps are run entirely separately, on separate servers, and neither code contains user data.”

She also said, “TikTok has never provided user data to the Chinese government, nor would we if asked.”

TikTok, whose chief executive lives in Singapore, hired a raft of well-connected American and European executives and security experts as political pressure intensified under Mr. Trump. It says it has no formal headquarters. It has sought to soothe American concerns by storing user data in the United States, with a backup in Singapore.

The American government’s security concerns come in two forms. The first, as Mr. Trump suggested in his executive order, is whether the vast trove of data TikTok holds — about the private sexual desires of fans of the app who might end up becoming American public officials, for instance — should be viewed as a national security issue. There’s no evidence the data has ever been used that way, and TikTok is hardly the only place Americans share details of their lives on social media. The second concern is whether TikTok censors politically sensitive posts.

A report this year by Citizen Lab, the cybersecurity watchdog organization in Toronto, suggested that both of these concerns are, at best, latent: It did not find any indication that TikTok was either censoring sensitive topics or transmitting data to China.

But TikTok’s glimpses of people’s inner lives are unusual. Another screenshot shared with me indicates that its content moderators have access not just to videos posted publicly, but also to content sent to friends or uploaded to the system but not shared, a difference from apps like WhatsApp and Signal that provide end-to-end encryption.

The second question is whether the Chinese government could use the platform to spread propaganda. After getting caught censoring a video condemning the mass detention of minority Muslims in China, TikTok has allowed criticism of the country's government. For instance, the hashtag #whereispengshuai, a reference to the Chinese tennis star who accused a top Chinese leader of sexual assault, autocompletes in the system, though TikTok videos with that hashtag have few views. There is no independent way of telling whether the company is suppressing the search, which has far more engagement on Twitter but similarly little on Instagram.

Some American analysts see TikTok as a profound threat; others view it as the kind of clueless panic that Americans now approaching middle age faced when their parents warned them that if they shared details of their lives on social media, they'd never get a job. Many, many other products, from social networks to banks and credit cards, collect more precise data on their users. If foreign security services wanted that data, they could probably find a way to buy it from the shadowy industry of data brokers.

"Freaking out about surveillance or censorship by TikTok is a distraction from the fact that these issues are so much bigger than any specific company or its Chinese ownership," said Samm Sacks, a cybersecurity policy fellow at the research organization New America. "Even if TikTok were American-owned, there is no law or regulation that prevents Beijing from buying its data on the open data broker market."

One thing that reporting this column has reminded me: The menace that TikTok poses to American national security appears to be entirely hypothetical, and depends on your analysis of both the U.S.-China relationship and the future of technology and culture. But the algorithm's grasp on what keeps me hooked — between trick tennis shots, Turkish food videos and all the other things it's figured out I like to watch — did pose a clear and present danger to my ability to finish this column.

Copyright 2021 The New York Times Company

Suggested sources

Looking for more sources on the same subject? Try these suggestions:

The New York Times

How
TikTok
Reads
Smith,
Ben.
**New
York
Times,
Late
Edition
(East
Coast);
New
York,**

EVALUATION AND
EMPIRICAL STUDY ON THE
INFORMATION SERVICE
Chen, Ping; Zhang,
Xiaojuan. **Eurasian
Journal of Social
Sciences;
Istanbul** Vol. 8, Iss. 2,
(2020): 53-69.