



INSTITUTE FOR THE STUDY OF CONTEMPORARY ANTISEMITISM

INDIANA UNIVERSITY

Robert A. and Sandra S. Borns Jewish Studies Program

Bloomington

Social Media & Hate Research Lab

TikTok's Archive for Researchers—A Tool in Name Only? An Experience Report

September 2024

As part of our work at the **Social Media & Hate Research Lab** at Indiana University, we've been using TikTok's Archive API from November 2023 to September 2024 to gather data for research into social media's impact on public discourse. While TikTok promises that their archive is a powerful resource for researchers, our experience has been anything but. The platform is plagued with technical issues, all our help requests have gone unanswered, and we've encountered persistent roadblocks that seem to make it deliberately difficult for researchers to use the tool effectively—despite TikTok's claims to the contrary. This is particularly concerning given the platform's stated commitment to supporting researchers and providing researchers access to their data.

1. Broken Pagination—Inconsistent and Unusable

According to TikTok's own documentation, their API should allow automatic pagination through up to 1000 requests without intervention. In reality, this has proven to be a false promise. Instead of automatic, smooth data collection, we have to manually reset the cursor after every 1-3 requests. Given that each request can retrieve up to 100 posts, this results in a time-consuming, frustrating process, where automation—central to any serious data collection—becomes nearly impossible.

There have been sporadic moments when the system worked better, allowing hundreds of requests to run automatically, but these instances are few and far between. At other times, the system barely functions at all, forcing us to stop after just a few requests and manually reset the cursor.

2. Missing Data—No Longer an Exception, but the Norm

Missing data is not a rare glitch in the TikTok Archive—it's an everyday occurrence that renders much of the collected information unreliable. TikTok has already acknowledged that its API may not return 100% of the data, but the situation has grown much worse. We regularly receive CSV files that are completely empty, or files that contain far fewer posts than the 100 allowed per request. This problem is no longer an occasional inconvenience but a constant barrier. For example, files that should contain 100 posts often contain only a handful, or none at all. This has become so prevalent that it's impossible to trust the integrity of the data.

Missing data and inconsistencies are also evident in the chronological order and time frames of the individual files, each containing up to 100 posts. In some cases, the time frames of the posts in these files overlap and when they do, they do not include the same posts.

How are researchers supposed to draw meaningful conclusions when the data we receive is so fragmented and inconsistent? It raises serious doubts about whether TikTok is intentionally limiting the information researchers can access—despite the platform's public claims of supporting research and transparency. It remains unclear whether the data are Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR).

3. Lack of Transparency—Is TikTok Truly Committed to Research?

Given the combination of broken pagination, frequent missing data, and inconsistencies, it's hard to shake the feeling that TikTok is not living up to its promises of supporting researchers. These issues do not seem like minor technical glitches but rather systemic problems that limit access to the platform's data. Researchers rely on data accuracy, consistency, and accessibility to carry out meaningful work.

Despite TikTok's public-facing promises, the company seems more interested in appearing transparent than actually delivering on that transparency. This raises serious questions about TikTok's true commitment to allowing independent researchers to explore and analyze its platform.

4. Demands for Improvement—TikTok Must Deliver on Its Promises

TikTok's archive has the potential to be an invaluable tool for understanding social media's impact, but that potential is far from realized. We, and many other researchers, need TikTok to:

- **Fix the broken pagination system** so researchers can retrieve data automatically, as promised.

- **Ensure complete and accurate data** rather than routinely providing empty or incomplete files.
- **Address and resolve inconsistencies** to guarantee that no data is systematically omitted.
- **Demonstrate genuine transparency** by supporting rather than hindering researchers' work.

If TikTok wants to be taken seriously as a platform that facilitates academic research, it needs to drastically improve its tools and deliver on its promises. Researchers cannot rely on a tool that creates more barriers than it removes. TikTok must act now to fix these issues and prove that it is committed to openness and research integrity. Only this will allow scholars to perform reliable research.

Contact:

Dr. Günther Jikeli, PI, Social Media & Hate Research Lab

Ph: (+1) 812 856 1150

gjikeli@iu.edu