

Thesis Project Portfolio

TikTok User Data: Determining the Critical Content

(Technical Report)

TikTok's Societal Impact of Data Collection

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Spencer J. Portuese

Spring, 2024

Department of Computer Science

Table of Contents

Executive Summary

TikTok User Data: Determining the Critical Content

TikTok's Societal Impact of Data Collection

Prospectus

Executive Summary

TikTok is an ever-growing social media that has the United States Congress worried about its data collection. Being a social media platform, it collects significant amounts of data on its users, which is amplified due to how core the algorithm of its “For You Page” is to its identity. In my technical research, I created a tool to parse the available user data to examine its content. In my STS paper, I use this tool to examine the data collected and analyze whether or not this significant data analysis is detrimental. I also discuss Congress’s concerns and why it would be a concern. Both papers are connected by examining both the technical aspects of what this data contains and then evaluating its impact on society. The impact of the data is left to be discussed in the STS paper.

The main reason the technical research was conducted was to be able to parse the large data file users can access from TikTok’s settings, and make it more clear what exactly it contains as it is hard to determine at a glance. This project was created by firstly examining data in the human-readable format, then once that was understood the application was made by parsing the more machine-readable format. Many fields were straightforward, such as IP address, birthday, and phone number, but most of the data is the timestamp that each video viewed was at, which required some additional parsing to display coherently.

The resulting application can take the machine-readable format of the TikTok data and return a PDF showcasing the pertinent information on each user, as well as heatmaps showcasing when videos are watched. The paper discusses potential further work that could be done, such as following links for videos watched and collecting the most frequently viewed hashtags to determine the potential interests of the user. The fields and data visualizations did not showcase anything specifically concerning or marked as a significant invasion of privacy outside of data users specifically provide to the app, such as their birthday.

Using this PDF, the STS paper asked the question of whether this data collection was beneficial to people using the app as well as impacting those outside of it by looking at the latent and manifest functions and dysfunctions of TikTok’s data collection. Some aspects of the data collection lead to positive results intentionally, but can also lead to negative results unintentionally, and in doing the data collection some negatives also happened intentionally. The paper analyzes all of these conditions to determine how data collection impacts society.

Looking at all of these consequences of data collection, it appeared that unless significantly more data is collected on each user which is kept secret from society, for most people the data collection tends to have a positive impact due to their experience on the app and the community it fosters, but the data collection can also affect how they share information in a negative way. However, that could be a big if, and “most people” does not include everyone. Congress’s concern is mostly about how China could use the data to influence or learn about persons of interest, which would harm those people and the American government. China technically could request TikTok’s company, ByteDance, to share all of their data with them, which could be used negatively against their adversaries. Furthermore, a larger concern with data

is children's use of the app and the information they put on the internet, which could potentially leave them vulnerable.

While there are many concerns over the data collection of TikTok, this research analyzed both the technical aspects of what the data contains and the societal implications that it can have. While users need to be careful what they share, most people generally can enjoy the app as-is without significant concern.