

During my Undergraduate Research Apprenticeship with Dr Sands, I collaborated on a research project examining how economic inequality is discussed and interpreted on TikTok. My work focused specifically on TikTok content relating to Luigi Mangione. His alleged assassination of Brian Thompson, the CEO of UnitedHealthcare, became a highly salient event that generated widespread online discussions about healthcare, wealth inequality, class conflict, and political violence.

I first conducted a literature review to understand how this incident had been studied on other social media platforms and to review existing research using TikTok data. This also involved familiarising myself with the TikTok Research API and understanding its capabilities and limitations for collecting large-scale social media data.

My first major task was to build the dataset used throughout the project. Using the TikTok Research API in R, I collected publicly available videos relating to Luigi Mangione, Brian Thompson, and UnitedHealthcare from 4 December 2024, the day of Brian Thompson's death, until 31 January 2025. This period captured the emergence and evolution of online discussions, memes, and reactions surrounding the incident. I extracted metadata for each video, including descriptions, hashtags, usernames, engagement metrics (views, likes, comments, and shares), and speech-to-text transcripts. Using the extracted video IDs and usernames, I generated TikTok URLs for each video before downloading the video files and associated metadata using yt-dlp.

The next part of the project focused on identifying the main themes. Initially, I applied BERTopic topic modelling to video descriptions and speech-to-text captions to explore common topics and identify broad themes. This provided an initial overview that informed the development of a manual coding framework. After reviewing a large sample of videos, we refined the codebook to include the following categories: relevance, crime details and legal proceedings, structural inequality (class and wealth inequality), calls to action, support for Luigi Mangione, support for the crime, and condemnation of violence.

I developed a stratified sampling approach based on both the date and an engagement measure that I created using views, likes, comments, and shares. Working alongside the other research assistants, I compiled a manually coded dataset that will ultimately be used to train and evaluate a Large Multimodal Model (LMM) capable of automatically classifying the full dataset of approximately 62,000 TikTok videos.

Throughout the project, I primarily used Python for data cleaning, organisation, and preparation. This included filtering metadata, extracting video IDs, identifying duplicate videos, matching downloaded video files to their corresponding metadata, and tracking videos that failed to download because they had been classified as restricted by TikTok. To maximise the completeness of the dataset, I also used the Chrome extension myfaveTT to download restricted videos manually and then matched these downloads back to the original metadata using the TikTok video IDs.

Following completion of the coding framework, we independently coded a shared sample of videos. I used Python to merge the individual coding files, standardise coding values, prepare the combined dataset, and calculate inter-coder reliability. I assessed agreement

between individual coder pairs using Cohen's kappa and calculated Fleiss' kappa to evaluate overall agreement across multiple coders before producing the final labelled dataset for machine learning.

Overall, the project produced a large, manually annotated dataset of TikTok videos relating to Luigi Mangione and broader discussions of economic inequality. This labelled dataset provides the training and testing data required for the Large Multimodal Model to classify the remaining videos automatically.

More broadly, the research examines how users engage with inequality-related content on TikTok and the types of frames and narratives that gain traction following highly salient political events. By combining manual content analysis with computational methods and large-scale machine learning, the project contributes to an emerging area of research on political communication within short-form video platforms, where comparatively little work has been conducted relative to platforms such as X (formerly Twitter). In particular, the project investigates how discussions of structural inequality intersect with support for, or condemnation of, political violence. The findings will contribute to ongoing policy debates surrounding platform governance, including the moderation of content that appears to support or normalise violence. The research aims to highlight emerging challenges for regulators seeking to balance freedom of expression and public safety in discussions of violence.