**Appendix A Data**

**A.1 Keyword Selection**

To ensure comprehensive data collection, we carefully selected a list of keywords that were used in our search query. These keywords were derived from fact-checking sources such as the China Internet Joint Rumor Platform (https://www.piyao.org.cn) and the International News Fact-checking Network (https://chinafactcheck.com/). We identified key topics and misinformation trends related to COVID-19 during the Omicron wave and utilized them to structure our dataset. Below is the list of keywords used in the search:

*Keywords used in the keyword search*

|  |  |
| --- | --- |
| Chinese | English |
| 新冠 连花清瘟 | COVID-19, Lianhua Qingwen |
| 新冠 北强南弱 | COVID-19, Strong in the north and weak in the south |
| 新冠 二次感染 | COVID-19, Double infections |
| 早阳早好 | The earlier getting positive the better |
| 转阴后就能放飞自我吗 | Can one be free after turning negative for COVID-19? |
| 复阳 | Tested positive again |
| 黄桃罐头 | Canned yellow peaches |
| 颜值高的人更不易得新冠 | Good-looking people are less likely to get COVID-19 |
| 奥密克戎小艾滋 | Omicron is “little HIV” |
| 奥密克戎后遗症 | Post-Omicron sequelae |
| 新冠 生育 | COVID-19, Fertility |
| 维c预防新冠 | Vitamin C for COVID-19 prevention |
| 白酒预防新冠 | Wine for COVID-19 prevention |
| 生理盐水预防新冠 | Normal saline for COVID-19 prevention |
| 美国 新冠检测 | COVID-19 testing in the US |
| 美国 核酸检测 | Nucleic acid testing in the US |
| 新冠大流行 | COVID-19 pandemic |
| 布洛芬 新冠 | Ibuprofen, COVID-19 |
| 抗原线条 | Stripes on the antigen test kit |
| 橘子汁抗原 | Orange juice antigen |
| 肺炎确诊 | Diagnosis of pneumonia |
| 新冠 电解质水 | COVID-19, Electrolyte water |
| 肺结节 疫苗 | Pulmonary nodules, Vaccine |
| 新冠 气溶胶 | COVID-19, Aerosol transmission |
| 新冠 终身免疫 | COVID-19, Life-long immunity |
| 新冠疫苗副作用 | Side effects of COVID-19 vaccines |
| 新冠疫苗保护期 | Protection period of COVID-19 vaccines |
| 新冠疫苗 致癌 | COVID-19 vaccines, Causing cancer |
| 新冠疫苗 变异 | COVID-19 vaccines, Mutation |
| 新冠 摘口罩 | COVID-19, Removing masks |
| 新冠 吃药顺序 | COVID-19, Order of taking medication |
| 新冠 药 | COVID-19, Medications |
| 新冠 出门 | COVID-19, Going outdoors |
| 新冠 肺癌 | COVID-19, Lung cancer |
| 新冠 大蒜 | COVID-19, Garlic |
| 新冠 制氧机 | COVID-19, Oxygen concentrator |
| 新冠 不良反应 | COVID-19, Adverse reactions |
| 新冠 痊愈 | COVID-19, Recovery |
| 新冠 美国 | COVID-19, USA |
| 新冠 日本 | COVID-19, Japan |
| 德特里克堡实验室 | Fort Detrick Laboratory |
| 新冠 特效药 | COVID-19, Effective drugs |
| 新冠 辉瑞疫苗 | COVID-19, Pfizer vaccine |
| 新冠疫苗 | COVID-19 vaccine |
| 新冠疫苗进口 | Import of COVID-19 vaccines |
| 辉瑞特效药 | Pfizer effective drugs |
| 西方躺平 | Western world lying flat |
| 欧美躺平 | Europe and America lying flat |
| Xbb | Xbb |
| Paxlovid | Paxlovid |
| 新冠 莫德纳 | COVID-19, Moderna vaccine |
| 新冠 阴谋 | COVID-19, Conspiracy theories |
| 新冠 源头 | COVID-19, Origin |
| 新冠 解封 | COVID-19, Lifting lockdowns |
| 新冠 放开 | COVID-19, Removing restrictions |
| 新冠 政策 | COVID-19, Policies |
| 新冠 医保 | COVID-19, Medical insurance |
| 新冠 审批 | COVID-19, Reviewing and approval |
| 新冠 出境 | COVID-19, Leaving the country |
| 新冠 入境 | COVID-19, Entering the country |
| 新冠 管理 | COVID-19, Management |

**A.2 Data Collection**

This study presents Tikcron (TikTok short video about COVID-19 during the Omicron wave dataset), a dataset covering the Omicron wave of the COVID-19 pandemic starting from late 2022 in China, collected from Douyin (Chinese TikTok). The dataset was constructed following these steps:

**Keyword Selection**. Fact-checking sources, including the China Internet Joint Rumor Platform and the International News Fact-checking Network, were used to identify relevant COVID-19-related misinformation and trending discussions.

**Video Collection.** We used Python to perform keyword searches on Douyin, retrieving videos related to selected keywords such as “新冠 二次感染” (COVID-19, double infections), “新冠 北强南弱” (COVID-19, strong in the north and weak in the south), “新冠 辉瑞疫苗” (COVID-19, Pfizer vaccine), and “新冠 美国” (COVID-19, USA).

**Metadata Collection.** We collected video-related metadata, including:

* Video attributes: Title, posting time, duration, number of likes, comments, and shares.
* Creator attributes: Username, verification status, introduction, and numbers of followings and followers.

**Timeframe.** From November 1, 2022 to March 31, 2023.

**Language.** The primary language of the dataset is Simplified Chinese (Mandarin).

**Filtering and Deduplication.** After removing duplicate content, the dataset was reduced from 128,294 videos to 42,201 unique videos.

**A.3 Data Accessibility**

The entire dataset used in this study is not publicly available due to platform restrictions and privacy concerns. However, metadata summaries and analysis results can be provided upon request for academic purposes. Researchers interested in accessing the data may contact the corresponding author for further inquiries.

**A.4 Ethical Approval**

This study was approved by the University of Southern California Institutional Review Board (IRB). The IRB designee reviewed the study’s application and attachments and issued the approval notice for the collection and analysis of data/specimens on February 22, 2023. The study ID is UP-23-00050.

This data collection process adheres to ethical guidelines, ensuring compliance with data privacy standards and avoiding personally identifiable information. Additionally, potential biases in dataset trends, such as topic prominence and demographic representation, are acknowledged as limitations of this study.