

Appendix B: Annotation codebook

Variables and coding instructions

This codebook provides guidance on the manual coding process applied to the X dataframe obtained by searching tweets obtained from the Community Notes data releases. The dataset contains 16 columns, each capturing specific numeric and categorical aspects of a tweet.

Tweet status

- **REMOVED:** A tweet is marked as “REMOVED” if it no longer exists or is inaccessible during the inspection.
- **NO-DATA:** A tweet is labelled as “NO-DATA” if it was posted before the rollout of impression data. This feature was progressively rolled out between November and December 2022, leading to inconsistent data cut-offs for impressions.
- **NO-MEDIA:** If a tweet does not contain any media.
- **NOT-AI-GEN:** If the media accompanying a tweet is not AI-generated, as determined by the consensus of community notes, the tweet is labelled “NOT-AI-GEN.”

User’s verified status (column name: verified)

- **TRUE:** The user posting the tweet has a verification tick, including any colour of verification tick and premium tiers such as organisation verification.
- **FALSE:** The user does not have a verification tick.

Political status (column name: political)

- **POLITICAL:** Tweets explicitly or implicitly referring to politicians, governments, political parties, or political issues.
- **NON-POLITICAL:** Tweets that do not meet the above criteria.

Media type (column name: media)

- **IMAGE:** The tweet contains a media identified as an image
- **VIDEO:** The tweet contains a media identified as video. Videos that only contain audio are still categorised as videos.

Labelling workflow

- **Initial inspection:** Review each row to assess the status of the tweet. Label as “REMOVED” if the tweet has been removed or does not exist. If impression data is missing, label it as “NO-DATA.” If the media is missing, label the tweet as “NO-MEDIA.”
- **AI image detection:** Inspect the image accompanying each tweet to determine if it is AI-generated. If not, mark the status column as “NOT-AI-GEN.” This annotation is based on community notes consensus noted during the inspection process.
- **Verification check:** Examine the profile of the user who posted the tweet to determine their verification status. Update the verified column with “TRUE” or “FALSE” accordingly.
- **Political nature:** Assess the content of the tweet to determine its political nature. Update the POLITICAL column with either “POLITICAL” or “NON-POLITICAL” based on this evaluation.