Ethics guidance around the use of Social Media data for dissertations (Hons and MSc Students)

These notes should be considered when filling out the MACS Ethics Form when using Social Media (Twitter (X), Instagram, YouTube, Facebook, etc) data for UG or PG dissertation (projects).

Human data from social media can involve username, user ID, name, address, personal details, and sensitive data among others. Note this list is not exhaustive.

When you use Social Media data, it will involve the use of Human data most of the time. The data can be classified either as publicly or as non-publicly available.

Non-publicly available data

If the data is non-public, it is considered as private data. If you have to log into the social media site to obtain data, then it is not considered to be public data. You need to read the "terms of use" of the social media site on the use of private data of their users.

Publicly available data

If you are gathering the data without logging in and the data is available to the public, e.g. Tweets, then you need to ensure that you:

- Do not record user IDs and personal data of users (perform data anonymisation)
- Use only aggregated data (do not quote tweets or posts) when publishing and sharing
- Consider vulnerable users (e.g. children, political views from some countries)
- Take care if the data includes pictures or videos
- Check the terms and conditions of the social media site regarding use of the users' data
- Twitter (X) does not allow the scrapping of its services without their prior consent. Consider using the Twitter API

Download secondary social media datasets

If you are downloading publicly available social media datasets from repositories or websites (e.g. Kaggle), you must ensure that the dataset has prior ethical clearance.

Examples of prior ethical clearance are:

- Ethical clearance from an ethics committee
- Data used in a peer-reviewed journal, which shows prior ethical clearance

If the dataset has human data, you must ensure to perform data anonymisation.