

## **Data Collection**

While I am not still 100% sure of my chosen topic just yet, if I do indeed go with: Machine Learning Techniques to detect Political on the X platform, I will probably use the X API for my data collection.

Looking at the X API documentation, it has all the endpoints for me to gather all my data. I should be able to retrieve content from posts, comments and user profiles.

The provided filters should let me query specific keywords and hashtags relevant to my research.

I will probably be writing a Python application to call the API. Ideally, I will be pulling the data in JSON format and storing it in a database. This is clearly the easy part. The more complicated one should follow straight away: data cleaning. This will be necessary to transform and convert the data so it is ready to train the model on.

This should provide me with all the necessary data for my research. However, I'll still be looking on the internet to check if there are any current dataset available. This would only reinforce the data and therefore, the model.