### **Project: Impact Analysis of Twitter's Takeover on Mastodon's Popularity**

# Goal:

* Investigate how the change in Twitter's ownership influenced user behavior and network growth on Mastodon

## Sub-Goals

* User Growth Analysis
  + Compare rate of new user registrations on Mastodon 10 months before and after takeover
  + Timeframe we’re looking into:  
    01.01.2022 Beginning of Analysis  
    27.10.2022 Twitter Takeover  
    31.08.2023 End of Analysis
  + Analyze demographic changes in the user base, if possible.
* Engagement Trends, regarding xyz topic:
  + Study the change in posting frequency and user engagement.
  + Identify shifts in peak activity times or days.
* Content Shifts:
  + Analyze the change in popular topics and hashtags.
  + Perform sentiment analysis to understand user reactions to the Twitter takeover.
* Network Dynamics:
  + Examine how the user interaction network evolved.
  + Identify any new communities or clusters that formed post-takeover.

# To do

* Finish setting up github
* ✅ Decide on which hashtags to filter (both capitalised and lowercase)
  + ✅ ‘twitter’
  + ✅ ‘Gaza’
  + ✅ ‘israel’
  + ✅ ‘thanksgiving’
  + ✅ ‘Milei’
  + ✅ ‘spotify’
  + ✅ ‘argentina’
  + ✅ ‘MissUniverseThailand2023’
  + ‘I\’m\’m’
  + ✅ ‘Trump’
* ✅ Decide on timeframe to filter → The last 30 days
* Data exploration: what to to focus on for data exploration, analysis and visualisation  
  → see suggested next page   
    
  Who does what?

| What | Who |
| --- | --- |
|  |  |

# Time frame for Thursday and Friday

| When | What |
| --- | --- |
| Thursday  09:00 | Morning stand up/**login**   * Git pull * Set goals for data exploration * Who does what? |
| Thursday  14:00 | First **meeting** to  **present** **results** and*optionally refine goals* |
| Thursday late evening  /  Friday  09:00 | Second **meeting** to  **show** **final** **results** |
| Friday  11:00 | **Decide** and **finalise** : **what** do we **present** (from our results)  Optional to do some last minute additional work |
| Friday  1400 | Prepare **presentation** |

# Data Preprocessing

* Create separate df or merge df?
* Which columns to drop?
* Consider  
  # Convert 'created\_at' column to datetime

df\_milei['created\_at'] = pd.to\_datetime(df\_milei['created\_at'])

# Resample to get daily counts

daily\_counts = df\_milei.resample('D', on='created\_at').count()

* Clean the 'content' column by removing URLs, special characters, and converting text to lowercase for Sentiment Analysis

# 

# Areas to focus on for data analysis and visualisation

**Most popular Hashtags amongst sample**

* Compare hashtag counts between Twitter and Mastodon.

**Daily Posts Trend for Hashtag #...** (created\_at)

* Analyse when posts with hashtags were most active.
* Find reasons for this: e.g. is it due to events like elections?

**User Engagement Analysis (** 'replies\_count', 'reblogs\_count', and 'favourites\_count' )

* Check summary statistics
* Compare engagement across the 10 hashtags.

**Language Distribution** (language)

* Analyse which languages are used within these hashtags.
* Show the language distribution amongst hashtags.

**Content Analysis** (content, spoiler\_text, media\_attachments)

* Visualize the most common words in posts with these hashtags. (word cloud?)
* Analyse the nature of content e.g. text, media etc.

**User Demographics and Behavior (**account.\* followers\_count, following\_count, statuses\_count)

* Analyse the profile statistics and behavior of users participating in these hashtags.
* Check for total most active users
* Analyse/Identify if accounts are bots (account.bot) or locked (account.locked

**Sentiment Analysis** (content, tags, language)

* Preprocessing
  + Clean the 'content' column by removing URLs, special characters, and converting text to lowercase.
* Sentiment Analysis
  + TextBlob or VADER
  + Apply to the 'content' column to get sentiment scores.
  + Classify each post as positive, negative, or neutral based on the sentiment score.
* Aggregating Sentiment by Hashtag
  + Group the data by hashtags.
  + Calculate the average sentiment score for each hashtag.
* Visualization

**User Network Analysis** (mentions, account.username, account.id, reblog, tags)

* Data Preparation
  + Create a list of edges where each edge represents an interaction (mention or reblog) between two users.
* Network Construction:
  + Use a library like networkx in Python to create a graph.
  + Add nodes for each unique user.
  + Add edges based on the interactions extracted in the first step.
* Analysing the Network:
  + Calculate network metrics like degree centrality to identify influential users.
  + Optionally, filter the network to only include interactions related to the specific hashtags.
  + Analyse interactions between users.
  + Identify unique users and create nodes in the network.
* Visualization:
  + Use networkx or matplotlib to visualize the network.
  + Highlight different types of interactions (mentions, reblogs) using different colors or line styles.
  + Use node size to represent user influence (like followers count or centrality measure).
* Hashtag-Specific Networks:
  + Optionally, create separate network graphs for each hashtag to compare the interaction patterns between them.

# Presentation

1. Introduction (1-2 minutes)

* Overview of Mastodon:
  + what is Mastodon is
  + Emphasize its decentralized nature and differences from platforms like Twitter.
* Purpose of the Study:
  + Introduce project's aim

2. Methodology (3 minutes)

* Data Collection
  + Mention data collected from Mastodon (time frame, hashtags, columns used).
* Tools Used:
  + Python, Pandas, Streamlit, etc.
  + methods such as Sentiment Analysis, Network Analysis, etc.

3. Key Findings (4-5 minutes)

* Temporal Trends
* Language Distribution
* Sentiment Analysis.
* User Network Analysis
* Content Analysis

4. Conclusion (1 minute)

* Summarize Insight
* Problems encountered
* Future Directions

5. Q&A Session