

Introduction to social media analysis

Lorenzo Cima



UNIVERSITÀ DI PISA

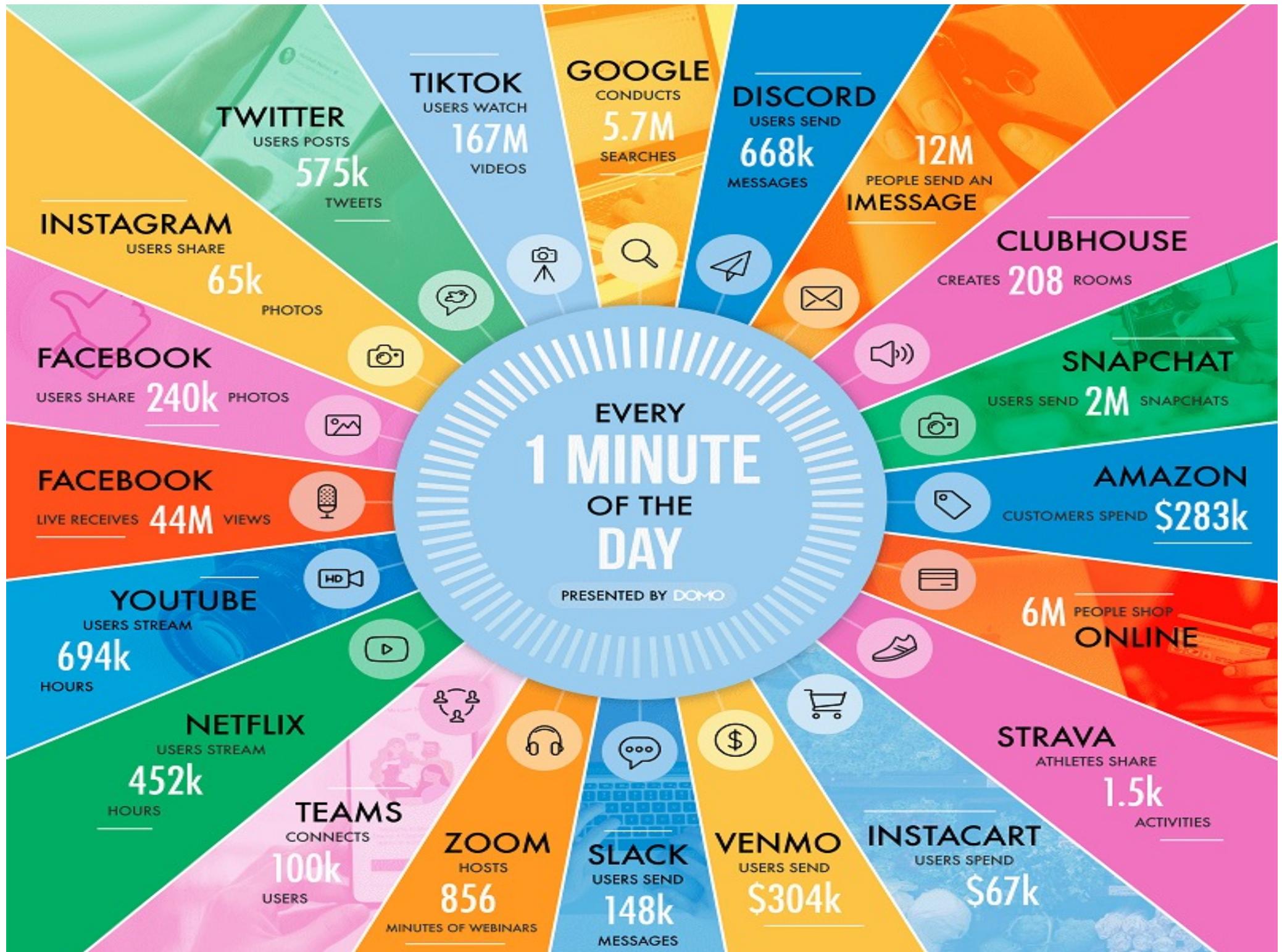


ISTITUTO
DI INFORMATICA
E TELEMATICA

lorenzo.cima@phd.unipi.it; lorenzo.cima@iit.cnr.it

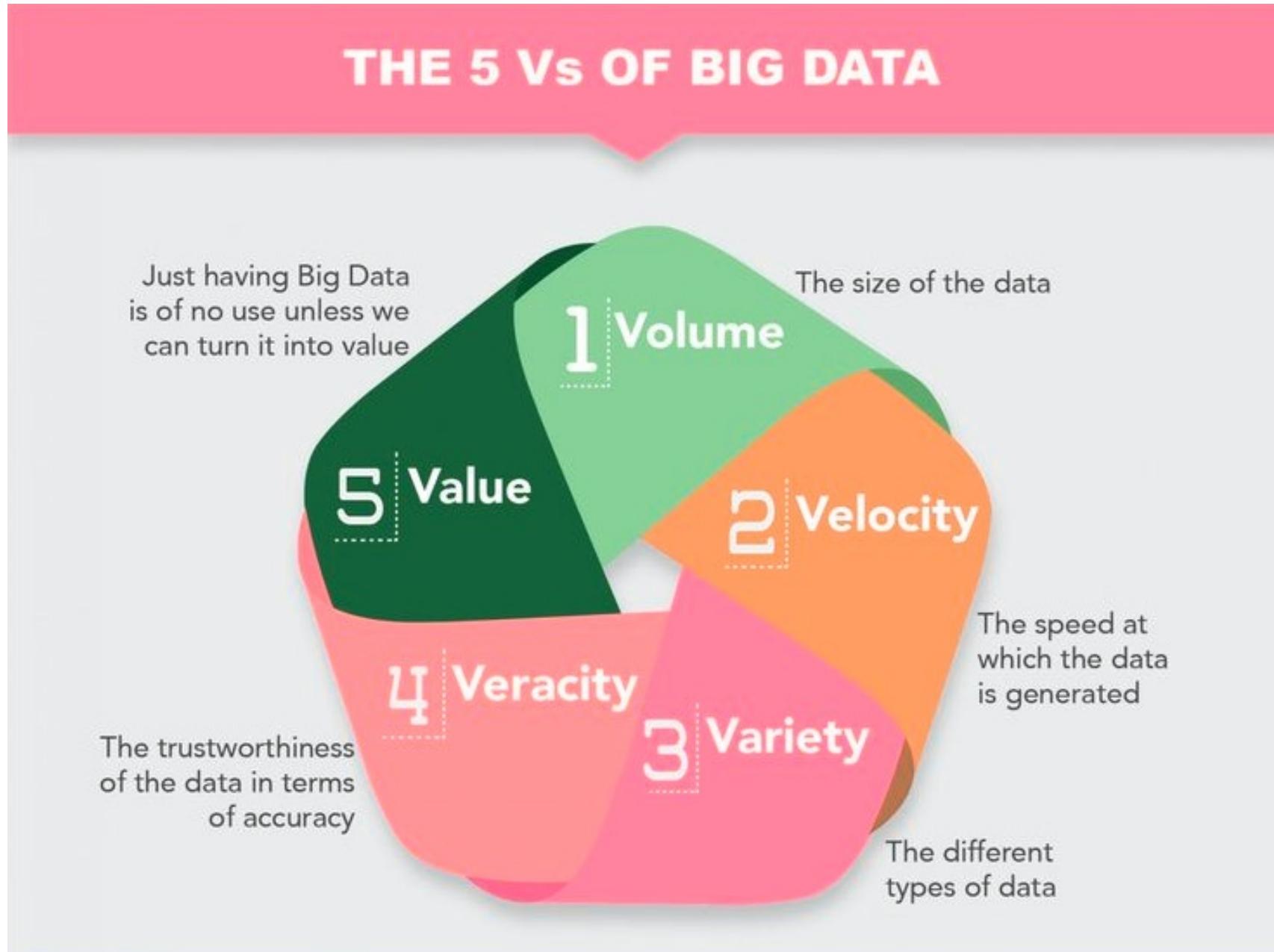


SOCIAL MEDIA ANALYSIS





Social Media = Big Data: The 5 Vs





Analyse Social Media

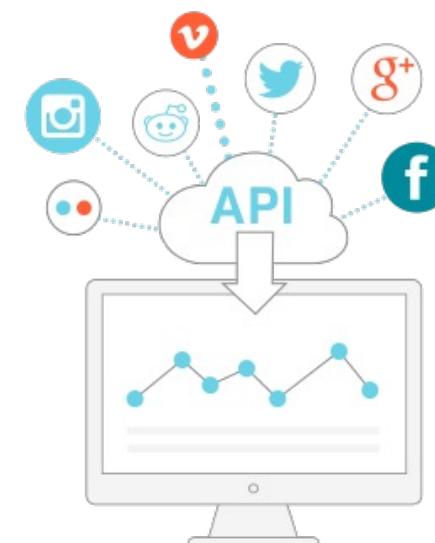
- Analysis tools



- Data resellers

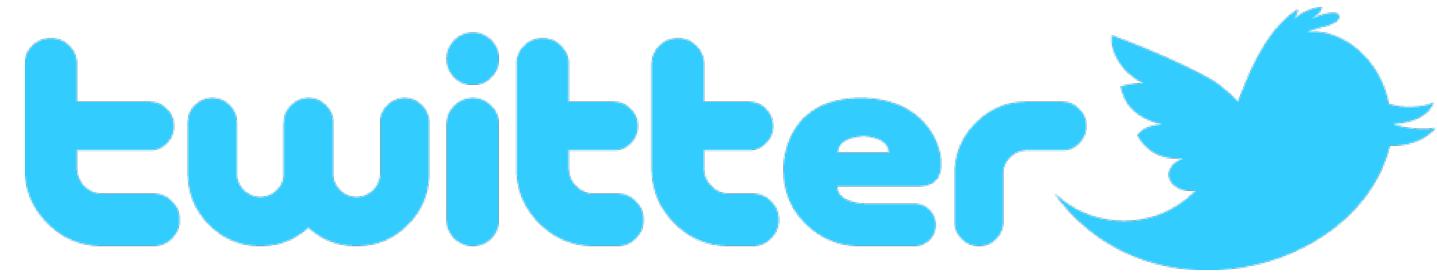


- Social media crawling





Analyse Social Media





Overview on Twitter

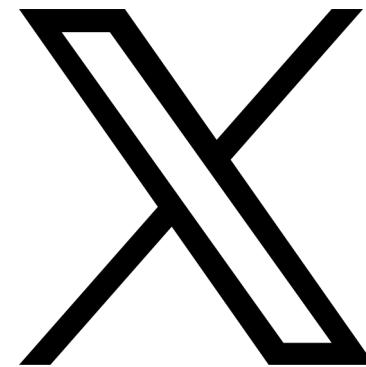


- free microblogging and social networking service
 - launched in March 2006
 - short text messages: max length is 280 characters (so-called *tweets*)
 - used to share news, as a participatory journalism tool
 - used by politicians, firms, and influencers as a direct communication channel to their followers
-
- **free data crawling for academic purposes (up to 10M tweets per month)**
 - **JSON data structure**





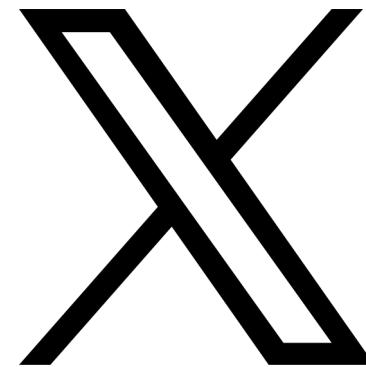
X: Twitter 2.0



- bought by Elon Musk in October 2022
- **main changes:**
 - paid verification
 - reinstatement of ~60k banned accounts
 - banned links to competing social media platforms



X: Twitter 2.0



- bought by Elon Musk in October 2022
- **results:**
 - more bots
 - more toxicity and hate speeches
 - less reliable information ecosystem
 - no free data for research purposes (\$5000 per month)
 - ...



Analyse Social Media

X



reddit



Overview on Reddit



- free social network, very popular in the USA
- launched in June 2005
- divided into topic-related communities, called *subreddits*
- on each subreddit people could upload a post (*submission*) or comment other submissions
- the information flow is controlled by *moderators*: volunteer who can take decisions if a submission/comment is inappropriate
- **Free APIs to retrieve data**
- **Presence of public dumps containing each public submission and comment**
- **JSON data structure**

**F A K E
C H A C T**

INFORMATION DISORDER

Malicious Accounts

Social media are fertile ground for the proliferation of malicious accounts

Why?

- Open platforms
- Anonymity
- Programmatic access (APIs)

For what purposes?

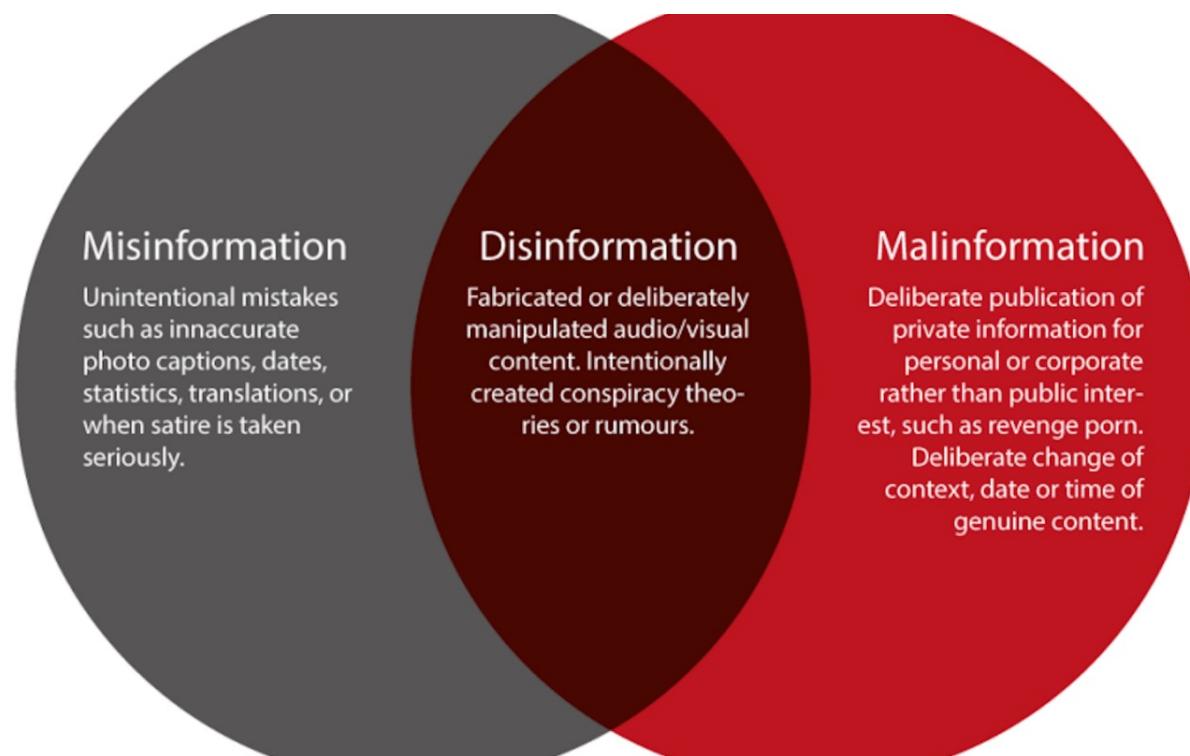
- Influence public opinions
- Economic purposes
- For fun





The Fake News Ecosystem

- misinformation (involuntarily fake or bad)
- malinformation (possibly true, but with malicious intent)
- disinformation (both fake and maliciously so)
- fake news (information at least partially untrue)





The Fake News Ecosystem

- hoaxes (deceiving information)
- rumors (unverified information)
- astroturfing (fake grassroots support for something/someone)
- propaganda (set of techniques for convincing someone of something)
- clickbait (sensationalist headlines/articles to attract readers)
- coordinated inauthentic/harmful behavior (colluded malicious actors)
- social bots (automated accounts, often deceiving and coordinated)
- trolls (colluded human-operated accounts, often coordinated)
- ...

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe.



Why did Fake News Become a Problem?

- fake news have always existed!
- but, social media **democratized** and **disintermediated** information production
 - anyone can create and spread content online
 - traditional news outlets lost their gatekeeping role

+

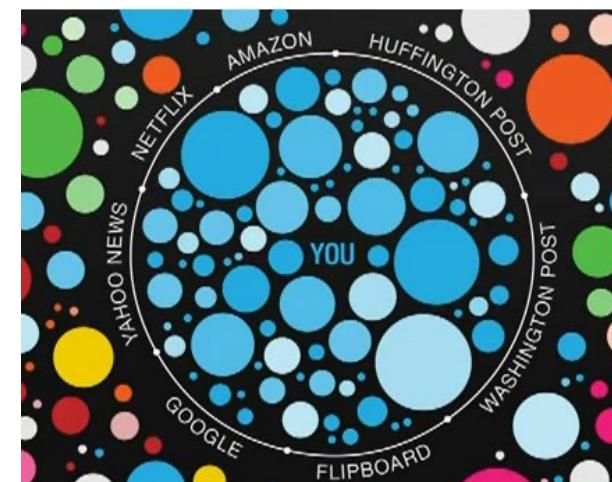
- politicians can directly talk to the electors
- traditional news outlets must fight for readers' attention

=

- less control over what is communicated and shared
- ... and provided means to **aggregate minorities**

Why did Fake News Become a Problem?

- we live in a time of information overload
- so, search engines and social media introduced **algorithmic filtering**, which led to **filter bubbles**



Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin UK.



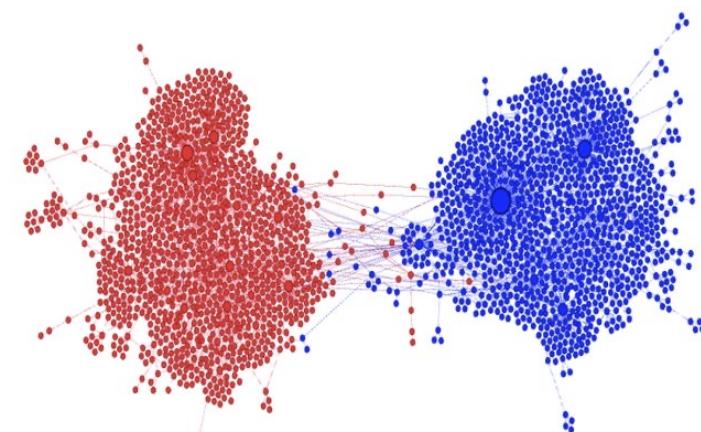
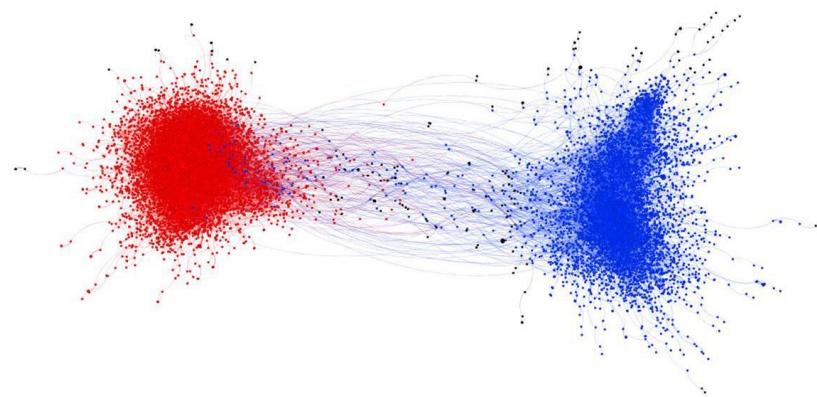
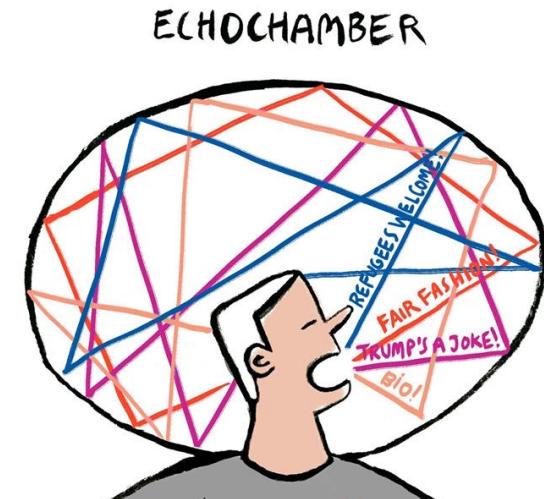
Why did Fake News Become a Problem?

- in addition, humans have a number of cognitive biases (e.g., **selective exposure/confirmation bias**)
 - tendency to favor and recall information that confirms our prior beliefs (also while searching for information)
 - tendency to interpret ambiguous evidence as supporting our existing position
- accepting information that challenges our own beliefs is tiring
- it is easier to accept information that confirms what we already believe
- we tend to stay in our comfort zone, also with respect to the information we consume

Cinelli, M., Brugnoli, E., Schmidt, A. L., Zollo, F., Quattrociocchi, W., & Scala, A. (2020). Selective exposure shapes the Facebook news diet. PLoS one, 15(3), e0229129.

Why did Fake News Become a Problem?

- filter bubbles and confirmation bias set the stage for the formation of **echo chambers**
- areas where like-minded people reinforce each other's opinions without being exposed to contrasting opinions
- **debunking** doesn't work inside echo chambers
- this further increases bias and **polarization**



Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. Available at SSRN 2795110.



Why did Fake News Become a Problem?

- the same also applies to many other moderation interventions
- some interventions **had no effects at all**:
 - contextual information to emphasize the source
- some interventions **had mixed effects**:
 - restrictions and bans of whole communities (e.g., Reddit's)
 - prompts against toxic posts (e.g., Twitter's)
- some interventions **backfired**:
 - friendship and content suggestions to reduce polarization
 - warning labels attached to debunked posts

Trujillo, A., & Cresci, S. (2022). Make Reddit Great Again: Assessing community effects of moderation interventions on r/The_Donald. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-28.



Why did Fake News Become a Problem?

- finally, social media also provide support for **anonymity** and easy ways to **automate** (via APIs) information production and diffusion:
 - **social bots**
 - trolls
 - cyborgs
 - ...
- **automation** is used to amplify effects and to reach visibility and large-scale impact

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications*, 9(1), 1-9.

Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49), 12435-12440.



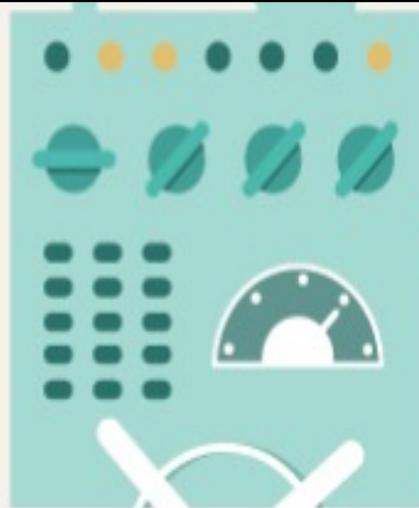
Why did Fake News Become a Problem?

- online **coordination** is also used to boost outreach and impact:
 - a single user cannot affect a whole social media
 - but a **large number of coordinated users** can!
- there exist **many forms** of online coordination, **not all of them are bad!**

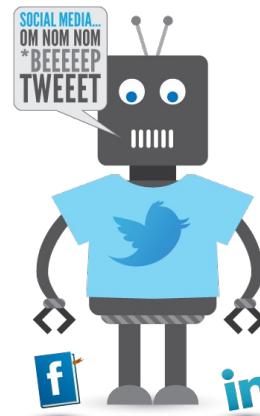
Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S., & Tesconi, M. (2021). Coordinated behavior on social media in 2019 UK general election. *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 443-454).



SOCIAL MEDIA ANALYSIS EXAMPLES



1. Bot Detection



2. Coordinated Behaviour



3. Content Moderation



Social Media Analysis





Social Bots

TWEETS 2,311 FOLLOWING 9,254 FOLLOWERS 8,572 LIKES 117

Tweets [Tweets & replies](#) [Photos & videos](#)

Oggi e per sempre, #metticilafaccia. NO alla mafia - 3 Nov 2015

Rai1 @RaiUno · 3 Nov 2015 Aspettando l'ultima puntata di #SottoCopertura, #metticilafaccia su Twitter. → ow.ly/TRHa8 #Rai1 #Rai

Contro la criminalità organizzata, twittate ora il vostro selfie con l'hashtag #METTICILAFACCIA.

SPAMBOT

2 Followers you know

Photos and videos

TWEETS 33.3K FOLLOWING 3,371 FOLLOWERS 4,039 LIKES 17K LISTS 4

Tweets [Tweets & replies](#) [Media](#)

Logopedista. Amo il mare ma sono ipertiroide.. ma chi se ne frega. Ci vado ugualmente. - 14m

Reggio Emilia Joined May 2013

[Tweet to](#)

3 Followers you know

1,912 Photos and videos

Italiani in fuga all'estero: nel 2015 espatriati in 107mila. Molti giovani quotidianonet.net/cronaca/italia... via @quotidianonet

UNIONE EUROPEA REPUBBLICA ITALIANA PASSAPORTO

Italiani in fuga all'estero: nel 2015 espatriati in 107mila. Molti giovani ... La meta preferita è la Germania. Lombardia e Veneto le principali regioni di emigrazione. Mattarella: "Cercare soluzioni" quotidianonet.net

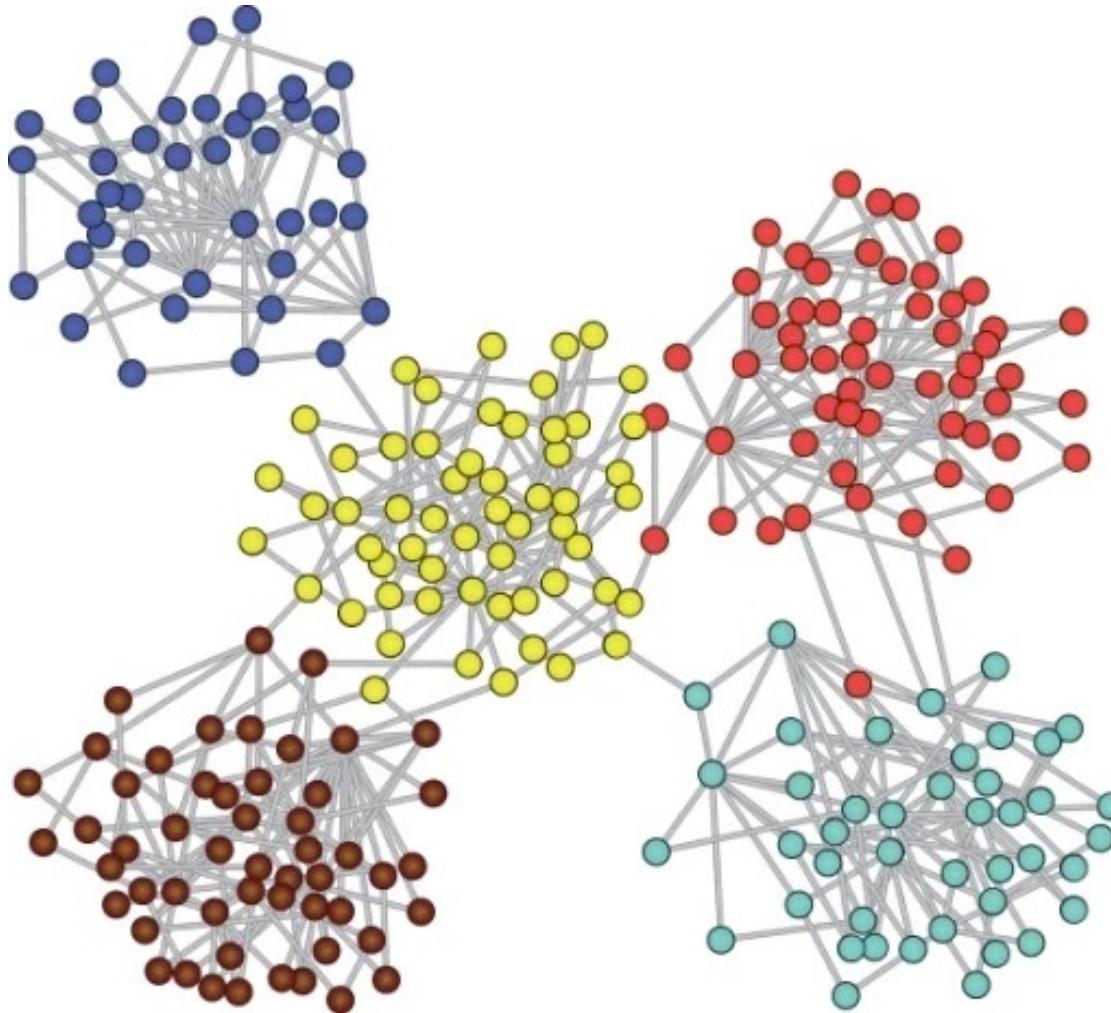
REAL



Botnets

The image displays a 4x4 grid of Twitter profiles, each featuring a placeholder profile picture of a cat. The users are: lorettaoari, Lori, annaticchi, pittore1960, Ross Ross, Verbenza, Carla Benni, and Viola Rossa. Each profile has a tweet from the account 'ValerioScanu' (@Valerio_Scanu) dated April 2018. The tweet reads: "Recita questo album è molto altro su Apple Music. Scopri la novità della tua musica preferita." Below the tweet is a link to a Spotify page for the album 'DIECI' by Valerio Scanu. The grid is set against a background of a blue-themed Twitter interface.

Social Media Analysis



COORDINATED BEHAVIOUR



Coordinated Online Behaviours

“

Groups of users performing synergic actions
in pursuit of a common intent

”

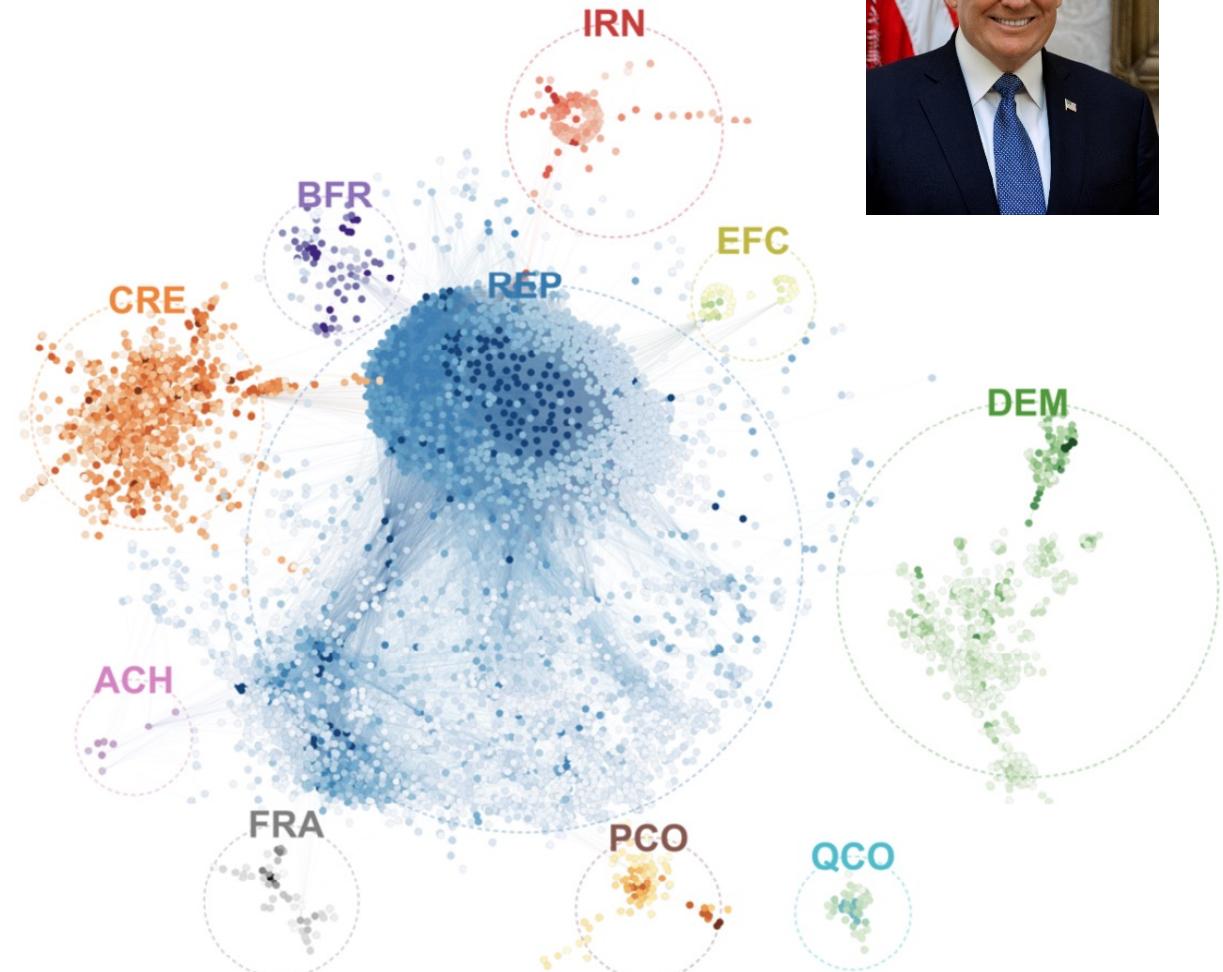




Detection of Coordination

- **DEM** and **REP**: main communities of democrats and republicans
- **CRE**, **PCO**, **QCO** and **EFC**: supporters of multiple conspiracy theories (pandemic, QAnon, election fraud)
- **IRN**, **BFR**, **FRA** and **ACH**: foreign Trump supporters

2020 USA presidential elections



Social Media Analysis



CONTENT MODERATION

Is disinformation the only problem of social media? **NO!**

- **Toxicity**
- **Hate speech and harassment**
- **Radicalization and polarization**
- ...



Need for appropriate countermeasures!

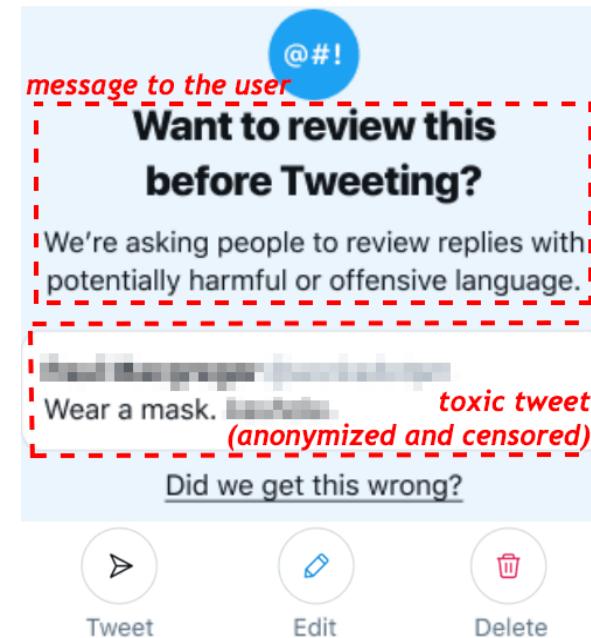


Countermeasure: Content Moderation



Soft interventions

- Warning messages
- Quarantine
- Visibility restrictions



Hard interventions

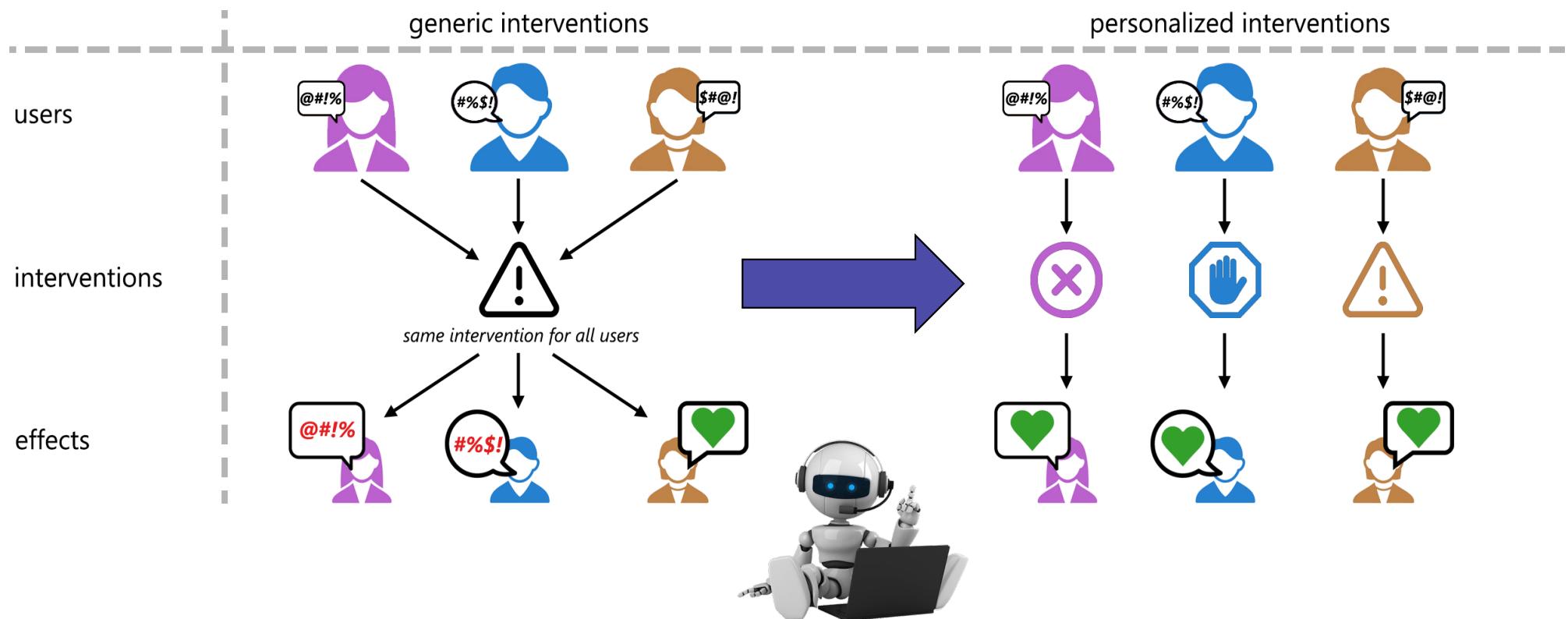
- Content deletion
- User deplatforming
- Massive bans



Content Moderation: Interventions

People react differently!

Need for personalized interventions



Usage of LLMs to develop automatic interventions



Requirements For Exercises

Python:

- version ≥ 3.8

Python libraries:

- networkx
- pandas
- scikit-learn
- statistics
- matplotlib



Gephi:

- <https://gephi.org/users/download/>

