Fighting Misinformation, Radicalization and Bias in Social Media

Erica Coppolillo[1,2], Carmela Comito[1], Marco Minici[1,3], Ettore Ritacco[4], Gianluigi Folino[1],
Francesco Sergio Pisani[1], Massimo Guarascio[1] and Giuseppe Manco[1]

[1] Institute for High Performance Computing and Networking, Italy
[2] University of Calabria, Italy
[3] University of Pisa, Italy
[4] University of Udine, Italy

Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
Terzo Convegno Nazionale
Pisa 29-31 Maggio 2023
cini National Lab AIIS

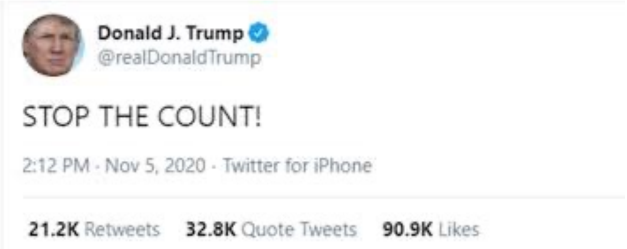
Outline

- Introduction
- Challenges
  - Fake News Detection
  - Radicalization in Recommender Systems
  - Bias and Fairness
- Conclusion

Social Media & Fake News

The spreading of misleading or fake news influences our society

Significant events were already seen in this decade



Donald J. Trump
@realDonaldTrump

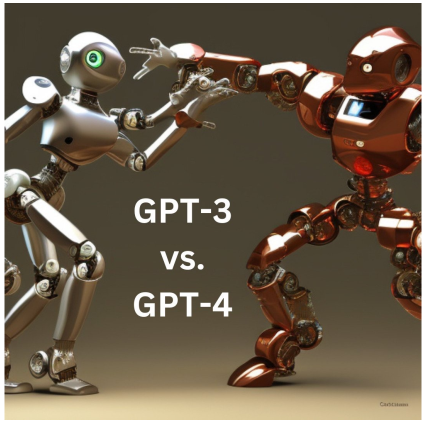
STOP THE COUNT!

2:12 PM · Nov 5, 2020 · Twitter for iPhone

21.2K Retweets 32.8K Quote Tweets 90.9K Likes

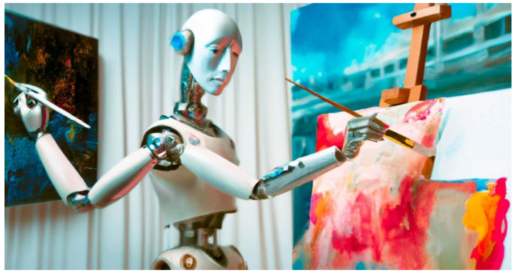
Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
Terzo Convegno Nazionale
Pisa 29-31 Maggio 2023
cini National Lab AIIS

Social Media & Fake News - Generation



Rapid technological advancements have made creating and spreading fake news remarkably easy.

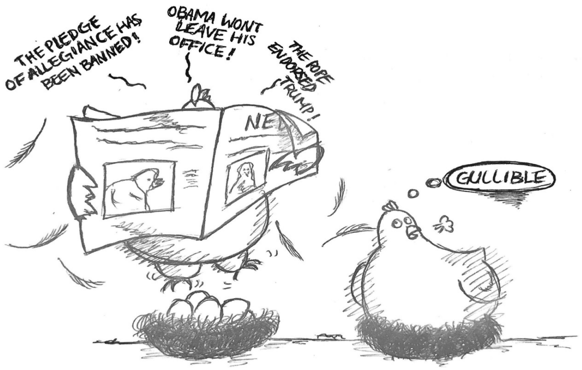
- **Text generation** tools like language models can produce seemingly authentic articles, blogs, and social media posts.
- **Image generation** tools employ deep learning algorithms to create realistic visuals that can deceive readers.



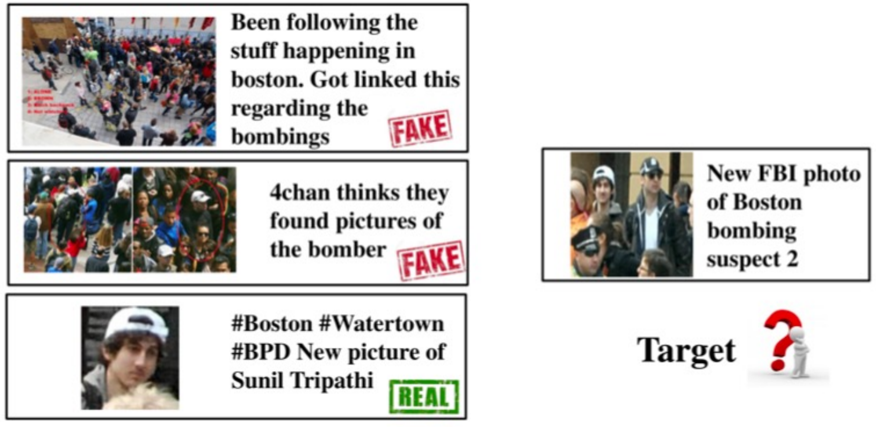
The combination of text and image generation tools amplifies the impact of fake news, leading to widespread dissemination and potential harm to public discourse.

Social Media & Fake News - Challenges

- The large use of social media has provided fertile soil for the emergence and fast spread of rumors
- Fake news is widely spread on social media and has detrimental societal effects
  - Fake news harms to real life


THE PLEDGE OF ALLEGIANCE HAS BEEN BANNED!
OBAMA WONT LEAVE HIS OFFICE!
THE POPE ENDORSED TRUMP!
NE
GULLIBLE

- *Uncovering fake news on social media is decisive and challenging*



Been following the stuff happening in boston. Got linked this regarding the bombings FAKE

4chan thinks they found pictures of the bomber FAKE

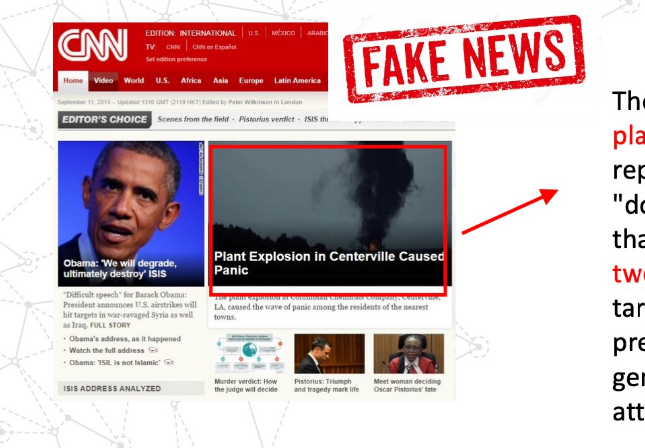
#Boston #Watertown #BPD New picture of Sunil Tripathi REAL

New FBI photo of Boston bombing suspect 2

Target

Fake News Detection Challenges

- Fake news are often generated on newly emerged (time-critical) events and domains and are hard to verify
- Fake news take advantage of multimedia contents to mislead readers and get rapid dissemination
- Fake news typically allow for sharing long text or short text


CNN
EDITION: INTERNATIONAL | U.S. | MÉXICO | ARABIC
TV: CNNi | CNN en Español
Set edition preference
Home Video World U.S. Africa Asia Europe Latin America
FAKE NEWS
EDITOR'S CHOICE
Scenes from the field · Pistorius verdict · ISIS th
ISIS ADDRESS ANALYZED

September 11, 2014 – Updated 1310 GMT (2110 HKT) Edited by Peter Wilkinson in London

Obama: 'We will degrade, ultimately destroy' ISIS

"Difficult speech" for Barack Obama: President announces U.S. airstrikes will hit targets in war-ravaged Syria as well as Iraq. FULL STORY

- Obama's address, as it happened
- Watch the full address
- Obama: 'ISIL is not Islamic'

Plant Explosion in Centerville Caused Panic

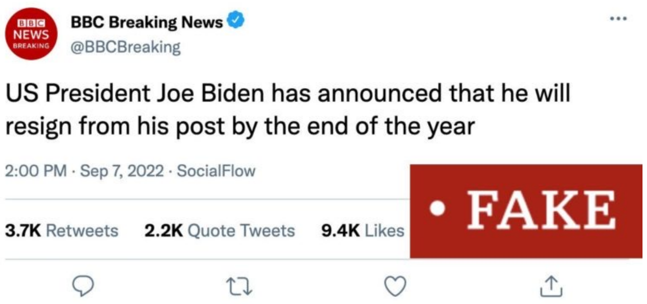
The plant explosion at Columbian Chemicals Company, Centerville, LA, caused the wave of panic among the residents of the nearest towns.

Murder verdict: How the judge will decide

Pistorius: Triumph and tragedy mark life

Meet woman deciding Oscar Pistorius' fate

The Columbian Chemicals plant explosion was reported to have involved "dozens of fake accounts that posted hundreds of tweets for hours, targeting a list of figures precisely chosen to generate maximum attention. "



BBC Breaking News
@BBCBreaking

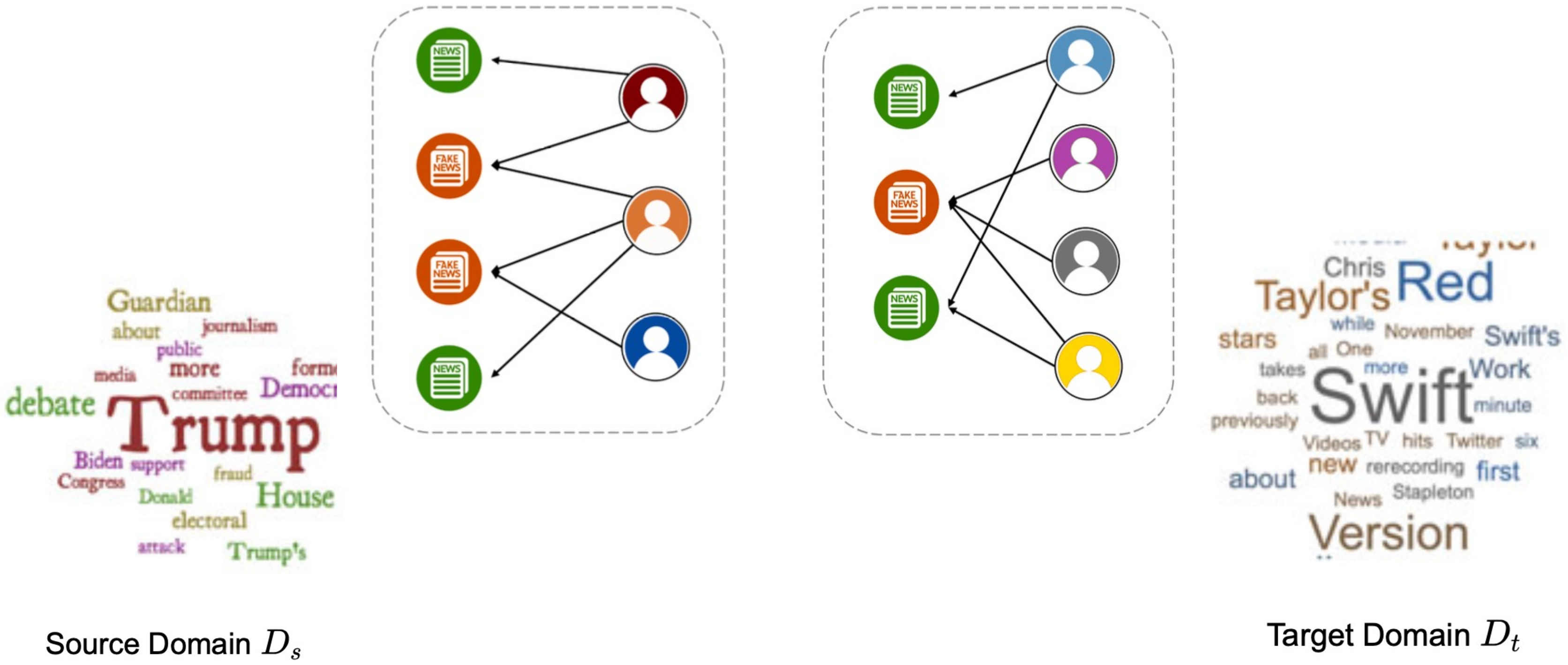
US President Joe Biden has announced that he will resign from his post by the end of the year

2:00 PM · Sep 7, 2022 · SocialFlow

3.7K Retweets 2.2K Quote Tweets 9.4K Likes

Fake News Cross-domain issues

- Domain-specific word usage
- Domain-specific propagation patterns
- Domain-specific community
  - users form groups containing people with similar interests (i.e., homophily)



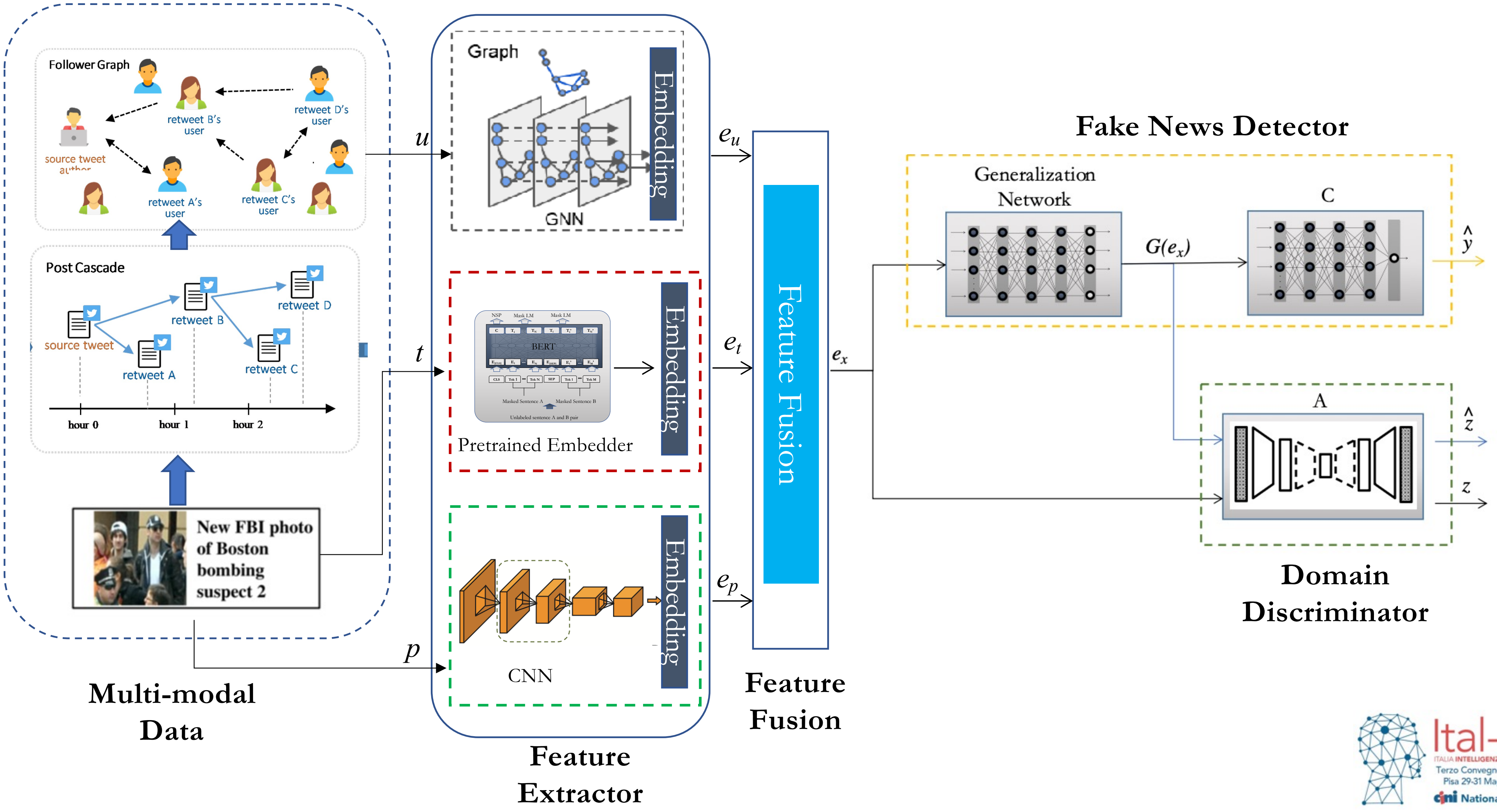
Fake News - Goal

- Central question: *“How and at which extent can the recent advancements in deep learning models together with social networks dynamics empower fake news detection?”*

- Investigate the use of deep learning models like *adversarial networks* and *autoencoders*
  - Learning multi-modal feature embeddings
    - Leveraging cross-modal information to preserve the relationships between different modalities
  - Generating domain-invariant feature representations to detect fake news on different domains and on new emergent events

- Objective: *Develop an innovative content verification tool, combining advanced deep learning techniques and multi-modal analytics technologies, to deliver solutions for news verification on social media and for general web content browsing.*

Fake News – Multi-Modal Framework

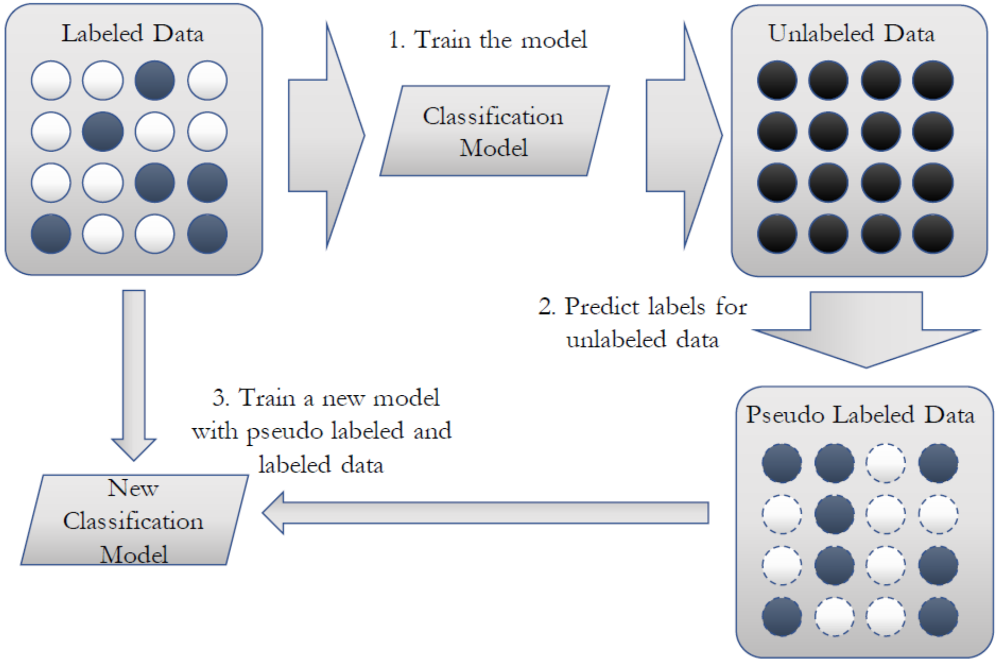


Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
Terzo Convegno Nazionale
Pisa 29-31 Maggio 2023
cini National Lab AIIS

Fake News – Training Strategy

Limited and unbalanced labeled data

Training with Pseudo-label strategy



Radicalization – Problem definition

- More focus should be put on how algorithms present content to users
- In recent years, the research community is questioning the **long-term effects** of recsys over users
  - They have been blamed for detrimental consequences such as *echo chambers*, *filter bubbles*, *polarization*, and *radicalization*
  - These phenomena refer to the tendency of enclosing users in a «**bubble**» and leading their preferences to the **extremes** (e.g., political leaning)

Radicalization - Real world Impact

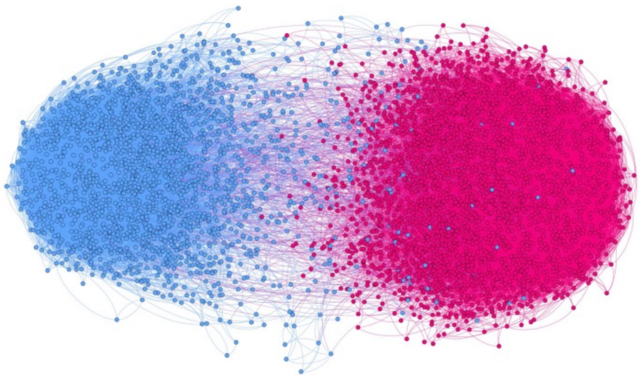
Some real-world **consequences**:

- It has been estimated that at least 800 people died and 5800 were admitted to hospital due to false information related to the COVID-19 pandemic, e.g., believing alcohol-based cleaning products are a cure for the virus
- A report estimated that over 1 million tweets were related to the fake news story “Pizzagate” by the end of the 2016 US presidential election*
- Recently, other crucial issues that require proper communication have started attracting public attention, like the Russia-Ukraine war

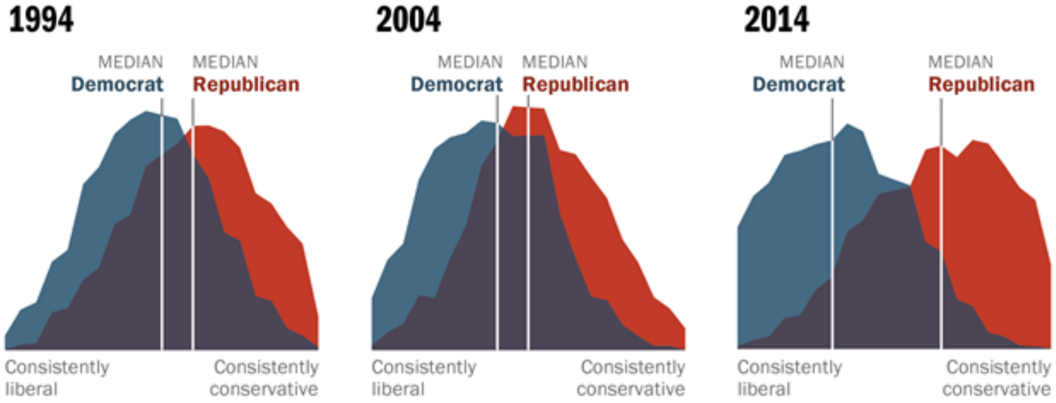
* https://www.bbc.com/news/blogs-trending-38156985

Radicalization & Echo Chambers

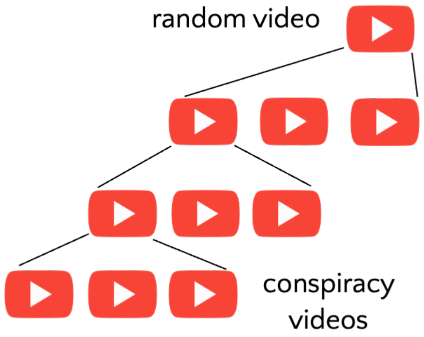
Social networks



Echo Chambers



Opinion Polarization



random video

conspiracy videos

Radicalization

Radicalization - Solution

Our contributions:

1. The definition of the **Algorithmic Drift** phenomenon and the introduction of **metrics** for evaluating it

2. A novel simulation **framework** for studying the long-term **influence** of any CF-based recommendation algorithm over users

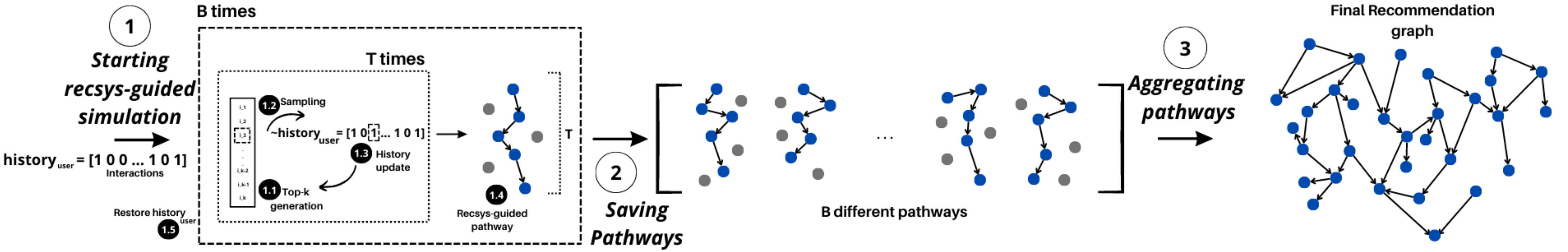
3. An extensive **analysis** showing that recsys can lead the user to deviate from their natural evolution

4. The formalization of the ***bridge effect***, by which a users sub-category does have impact in the influence dynamics of the whole population

Radicalization - Framework

Goals:

- Building a simulation framework to evaluate the effects of RecSys on the users’ preferences
- Investigating, quantifying and mitigating the **influence** of CF-based recsys in terms of *algorithmic drift*

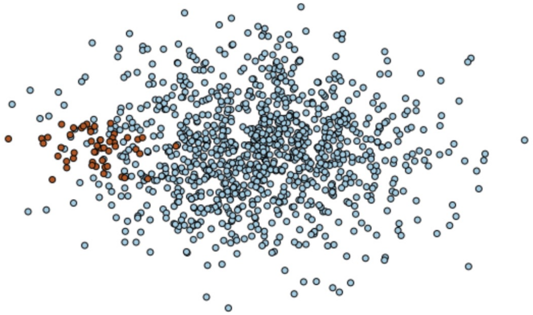


Mitigations → PRE-processing based strategy

Mitigations → POST-processing based strategy

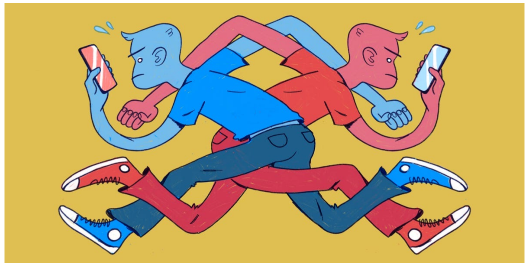
Bias & Fairness

- ML-based recommendations aim to personalize user experiences and enhance engagement by suggesting relevant items or content. However, these recommendation systems can suffer from inherent biases, leading to issues of fairness and discrimination.
- Bias can emerge due to several factors, such as biased training data, algorithmic design, and user feedback loops. They can perpetuate stereotypes, reinforce existing inequalities, and limit users' exposure to diverse perspectives.
- Fairness concerns arise when certain groups are systematically favoured or disadvantaged by the recommendations, based on factors like race, gender, or socio-economic status.



Bias & Fairness (2)

- Addressing the problem of bias and fairness requires careful algorithmic design, diverse and representative training data, and ongoing monitoring and evaluation
  - *We addressed this problem in the case of RecSys. We modeled a DL architecture that significantly improves the exposure of low-popular items vs medium-high ones*
- Ethical guidelines, transparency, and user control over recommendations are vital for promoting fairness and mitigating the negative consequences of biased ML-based recommendations



Conclusion

Addressing the consequences of fake news on social platforms and alleviating its effects require a multi-level approach involving:

- fact-checking mechanisms and critical thinking
- enhanced algorithmic transparency
- innovative solutions to combat the spread of fake news effectively
- responsible sharing, discouraging the dissemination of unverified news
- and much more 🙂

Ital-IA

ITALIA INTELLIGENZA ARTIFICIALE

Terzo Convegno Nazionale
Pisa 29-31 Maggio 2023

cini National Lab AIIS

Thank you for your attention!

francescosergio.pisani@icar.cnr.it