chist-era CHIST-ERA Call Development Conference 2021 - Topic 2 - OSNEM #7

Practice area

Click and drag the post it to move it

Practice adding notes to the MURAL board here

Drop the post it note here

Double click to produce a post it note and type on it

trial

Double click tag to type your name and affiliation, drag to your location on the map

Tag your name and affiliation

Name, Location

Name, Location

Name, Location

Name, Location

Name, Location

Name, Location

Name, Location

Name, Location

Zoom in and add post it notes

Free Parking

This space is for any ideas you may have that there is no place for, or any feedback you would like to give the organisers

Session I

Double click where you would like to write and strta typing

Brainstorming

Well Sorted Outputs

In the mid-term, new synthetic media generation approaches may be able to create synthetic media streams that consistently represent whole alternative realities. These pose new risks and call for new detection approaches.

The simple textual format of twitter allow bots to disguise as real people & engage in misbehavior. With virtual conversational agents & deep fakes, video-based misbehavior will be possible on OSNEM beyond twitter. How do we detect & address such attacks?

As artificially created intelligent agents are becoming more easy to create it is becoming crucial to be able to recognise such activities as they can easily be used for malicious purposes

Deciding whether user-generated content is violating certain ToS or community norms is a highly subjective/complex task. For this, more sophisticated approaches are needed for detecting harmful online behavior without putting freedom of speech at risk.

It is of interest to detect the use of AI-based techniques for the automatic creation and manipulation of visual data (e.g., GANs for image/video generation, motion transfer, video translation, style transfer, automated inpainting)

Once autonomous agent and media synthesis algorithms reach new capabilities, we expect complete services and suites for creating and managing online armies for different goals (propaganda). Detecting such activities will be crucial for online safety.

It is of interest to detect re-sharing of visual media content that has not been produced/acquired natively by the user but has been already shared through the same or other web services.

The issue with fake news is not much their existence, but their (excessive) circulation. How can we leverage an interdisciplinary approach to study these dynamics, including mathematicians, computer scientists, social scientists, engineers?

Plainly false content is clearly a bad thing, but at least we can try to identify it. Instead, there is a broad grey area of low-quality content that are hard to classify. Hence the need to understand and mitigate the circulation of these "fakish" news.

Synthetic media generation techniques are getting so good that it will soon be difficult to differentiate it from authentic content. Methods and workflows are urgently needed for detecting inauthentic media content

As online content has already exploded intelligent methods are becoming essential for the automatic identification of harmful content, ensuring that this is done reliably so that freedom of speech is not harmed

Misbehaviors are made possible and sometimes favored by the way the information is shared and attention is sought on the platforms. These dynamics can only change by chnging the platform design, which is consequence of economic incentives.

Overarching Themes

Double click where you would like to write and start typing

Overarching Theme

From the points made in Part 1, list the overarching themes that can be identified

1. Means and methods to quantify the effects/impacts, this can be offline or online behaviour within the platform i.e. is an intervention is successful
2. measurement of the change of behaviour overtime and from multiple sources (incl. online and offline)
3. blurred boundary between what is outright fake or low quality
4. understand vulnerability online - and the susceptibility to misinformation (age, mental health, cultural background, economics)
5. detecting misinformation in multimedia
6. detecting misinformation across multiplatform
7. detecting misinformation in multiple languages
8. Benefits for actors and platforms spreading misinformation

Double click the grey space to start writing

Session II

Double click the grey space to start writing

Ranking Overarching Themes

From the points made in Session I, rank and add a description to the Theme

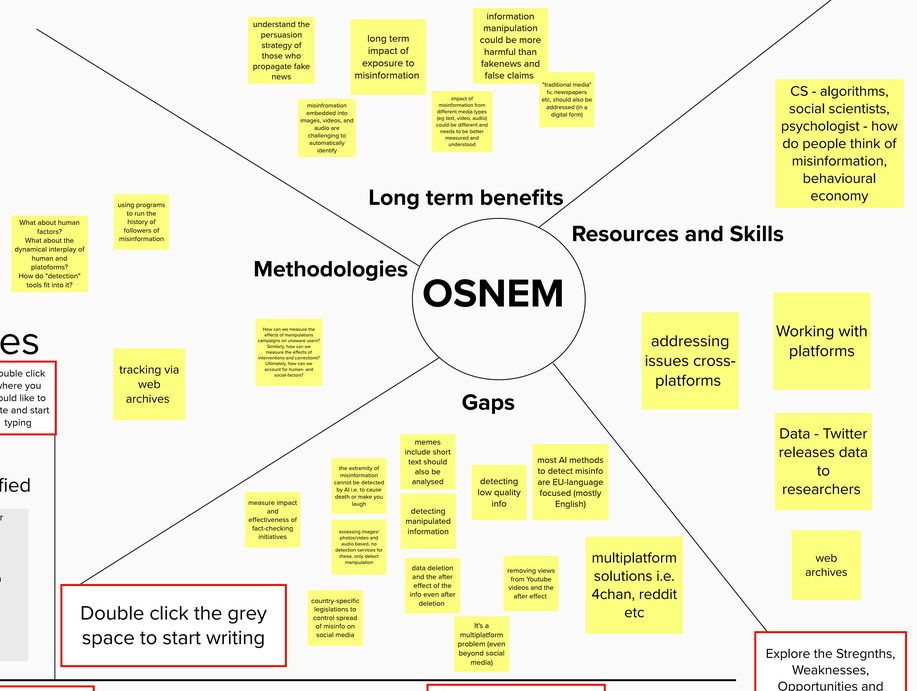
1. Means and methods to quantify the effects/ impacts, this can be offline or online behaviour within the platform i.e. is an intervention is successful
(Short and Long-term impact)

2. Evaluating vulnerability and susceptibility to misinformation (demographics, economics, education, Online/offline exposure, position within network). Interplay of interaction and feedback

3. Advancing methods for detection of misinformation (or otherwise low-quality content) in multimedia, across multiplatforms and multiple languages.

4. Investigating evolution of claims, motivation and impact covering time and location

Zoom in and double click to produce a post it note



Explore the Stregnths, Weaknesses, Opportunities and Threats for the Overarching Themes from the previous session

Zoom in and add post it notes

SWOT Analysis

Overarching Theme #1

Strength	Weakness
identify most efficient intervention techniques extends ongoing work	unclear how to detect/measure success of an intervention long-term studies require longer project lifespans
Opportunity	**Threat**
this step is a necessary prerequisite for many other analyses possibility to rank campaigns/manipulations based on their effects. Allows to prioritize interventions on most dangerous threats	possibility to measure no effects, not because of the actual lack of effects, but due to insensitive methodology or narrow scope of the analysis data from multiple platforms difficult to get

Overarching Theme

Strength	Weakness
novel area of research beyond just age and skills concept of overall exposure	requires measuring/ inferring many different factors (e.g., demographic factors, social and interaction networks, etc.)
Opportunity	**Threat**
better policies and protection measures understanding vulnerable groups and reasons behind that	could lead t- more targetted misinfo campaigns

Overarching Theme #3

Strength	Weakness
more and better detectors can allow us to better track the spread of mis/disinfo going beyond true/false detection	supervised methods need lots of high-quality training data: is this available? for which tasks? bad actors change their strategies over time: need to constantly adapt detectors bias, propaganda, sensationalisation of content, and many other parameters need to be researched
Opportunity	**Threat**
more fine-grained detectors can allow us to move from binary outcomes (e.g., true vs fake) to more nuanced ones, which would be more useful	the most powerful methods (deep learning based) are typically the least explainable and interpretable ones. (weakness considering implications) misuse of results by misinfo generators

Overarching Theme #4

Strength	Weakness
understand influence of multiple media going beyond single platforms	claims might reappear with no intentional connections motivation often unclear or not possible to determine automatically
Opportunity	**Threat**
better technical detection methods more convincing explanations	exposing financial benefits of spreading misino data from multiple platforms difficult to get

Session III

Zoom in and double click to produce a post it note

Maximising Success.....

How can Weaknesses and Threats be addressed

Refer back to the SWOT analysis and discuss how the weaknesses and threats could be mitigated

many threats and weaknesses can be addressed by having a multidisciplinary team would help, covering CS, social sciences

time needs to be taken into account for the mathematical models and should be sizable, in order for the data to have a temporal dimension

time is needed to measure effects mentioned in #1

expertise to work on different platforms

expertise in extracting data from different platform

missing data?? awareness of missing data is important

What is deliverable in a 2-3 year project

What could a CHIST-ERA Project potentially deliver?

a project could deliver methods i.e. effects of manipulation campaigns

methods for detection can be improved

less than 3 years cannot hire PhDs

2-3 years is a reasonable amount of time

opportunities to assess global events during the project lifetime

Transnational difficulties

What challenges could be faced by a BCI CHIST-ERA project due to its transnational nature

cultural differences plays a role in the spread of information, there will be different interpretations

pandemic? travel and collaborative working

dissemination of project results

language differences, different news outlets in different countries, may need multi lingual systems, may lock out other countries from your research and methods

ethics and laws across nations, different data sharing policies may cause issues