



chist-era



CHIST-ERA Projects Seminar 2023
*Foundations for Misbehaviour
Detection and Mitigation Strategies in
Online Social Networks and Media
(OSNEM)*

April 04, 2023



Programme co-funded by the
EUROPEAN UNION



- ◆ **CON-NET**: Integrating CONtent and NETwork structure for detecting, understanding, and mitigating online misbehaviour
 - BE, IRL, FI, *LU*, *TU*
- ◆ **HAMISON**: Holistic Analysis of Organised Misinformation Activity in Social Networks
 - CH, *EE*, **ES**, FR
- ◆ **iTRUST**: Interventions against Polarisation in Society for TRUSTworthy Social Media
 - BE, CH, SP, *PL*
- ◆ **MARTINI**: Malicious actors profiling and detection in Online Social Networks through Artificial Intelligence
 - *EE*, **ES**, FR, *LT*, *TU*



Objectives of the call

- ❖ Create a **foundation** for identifying peculiar behaviours of **malicious** entities
- ❖ Create the basis for **detection and mitigation strategies** for misbehaviour in OSNEMs
- ❖ Provide societal change towards OSNEM **literacy** and citizen **participation**
 - ✓ Detecting malicious actors (e.g. bots and trolls)
 - ✓ Identifying low-credibility information and fake news
 - ✓ Increasing the users' awareness and mitigating misbehaviours



Detecting malicious actors

Human actors vs non-human actors (bots)

Looking at **network** and **content** characteristics

Approaches:

- Social Network Analysis
- Community Detection
- Natural Language Processing
- Multilayer Visual Analytics
- Profiling malicious actors behaviour



Identifying low-credibility information and fake news

At the **content** level - multilingual and multimodal

Approaches:

- Modelling disinformation **narratives** and **intentions**
- Leveraging **rhetorical strategies** from linguistics and philosophy: ethos/pathos
- Development of deep learning models (e.g. Transformers, GANs, Siamese architectures, etc.) to detect fake news in **multimodal** content



Increasing the users' awareness and mitigating misbehaviours

Supporting (non)-professional stakeholders

Outcomes/approaches:

- Providing a **taxonomy of malicious actors** as an effective tool for pattern recognition
- Designing **intervention** methods for mitigating disinformation spread through **reframing**
- Observing and evaluating **intervention effects** using agent-based modelling
- Delivering **visual analytics** platform for understanding misinformation
- Conducting qualitative studies with stakeholders to understand **exposure** and subjective **experience** of misinformation



Use Cases

- Journalists understanding exposure to misinformation
- Fact-checkers early detection of viral propagation of misinformation
- Analysts international observers for political elections
- Debating teams finding errors in their argumentation
- Epidemiology monitoring public health
- Social sciences understand users behaviour from interdisciplinarity perspective

Application domains

- Anti-vaccination [iTRUST, MARTINI, HAMISON]
- Anti-immigration [MARTINI, HAMISON]
- Climate change [iTRUST, HAMISON]
- Political elections (national and European Level) [CON-NET, HAMISON, MARTINI]
- COVID-19 [CON-NET, MARTINI]



EC Digital Strategy: Shaping Europe's digital future

- Policy on tackling online disinformation

Calls exist in topics adjacent, but none are OSNEM focused

Pillar II: Global Challenges

- Cluster 2: **“Culture, Creativity and Inclusive society”**
 - Strengthen European democratic values
 - HORIZON-CL2-2024-DEMOCRACY-01 - Computational Social Science approaches in research on democracy
- Cluster 3: **“Civil security for society”**
 - Misinformation as a threat
- Cluster 4: **Digital, Industry and Space - Research and innovation:**
 - Shaping trusted technologies
 - HORIZON-CL4-2021-HUMAN-01-27: AI to fight disinformation



Fostering RRI through criteria and guidelines:

- ❖ Ethics and research integrity
 - Bias in the collections
 - Maximising anonymization while enabling research
 - Impact on subjects and researchers
 - interdisciplinarity
- ❖ Gender equality and diversity
 - Balance in teams
 - Balance in data collection
 - Balance in stakeholders
- ❖ Open Access/Open Science
 - Partnerships with non-profit organisations and industry
- ❖ Public Engagement
 - Collaboration with non-profit organisations
- ❖ Science Education
 - Education partners working with young people



chist-era

Open Science

- ❖ Open-access publications
- ❖ New datasets
 - ✓ GDPR restrictions vs Reproducibility and data sharing
- ❖ Dissemination seminars, workshops
 - ✓ Call topic seminar with all projects (Dagstuhl proposal)
- ❖ Surveys, tutorials, roadmaps
- ❖ Open, public technical reports





chist-era

Technology Transfer

- ❖ Most project results open source / permissive license where possible
- ❖ Exploitation ready results may be protected
 - ✓ Ownership follows inventorship

Main targets for transfer

- ❖ Fact-Checkers
- ❖ Journalists
- ❖ Analysts
- ❖ Young people; student debating teams

Development of demos and tools

Involvement of industry

Long term obstacles:

Real integration on social media platforms



Questions ?