# CHIST-ERA Projects Seminar 2021
## *Topic XAI*

**Raphaël Troncy – CIMPLE**

**raphael.troncy@eurecom.fr / @rtroncy**

**April 13, 2021**

# CIMPLE: Countering Creative Information Manipulation with Explainable AI

April 2021 - March 2024, 1.1M€ funding
**https://www.chistera.eu/projects/cimple**
**https://www.cimple.eu/** **(to be launched this month)**
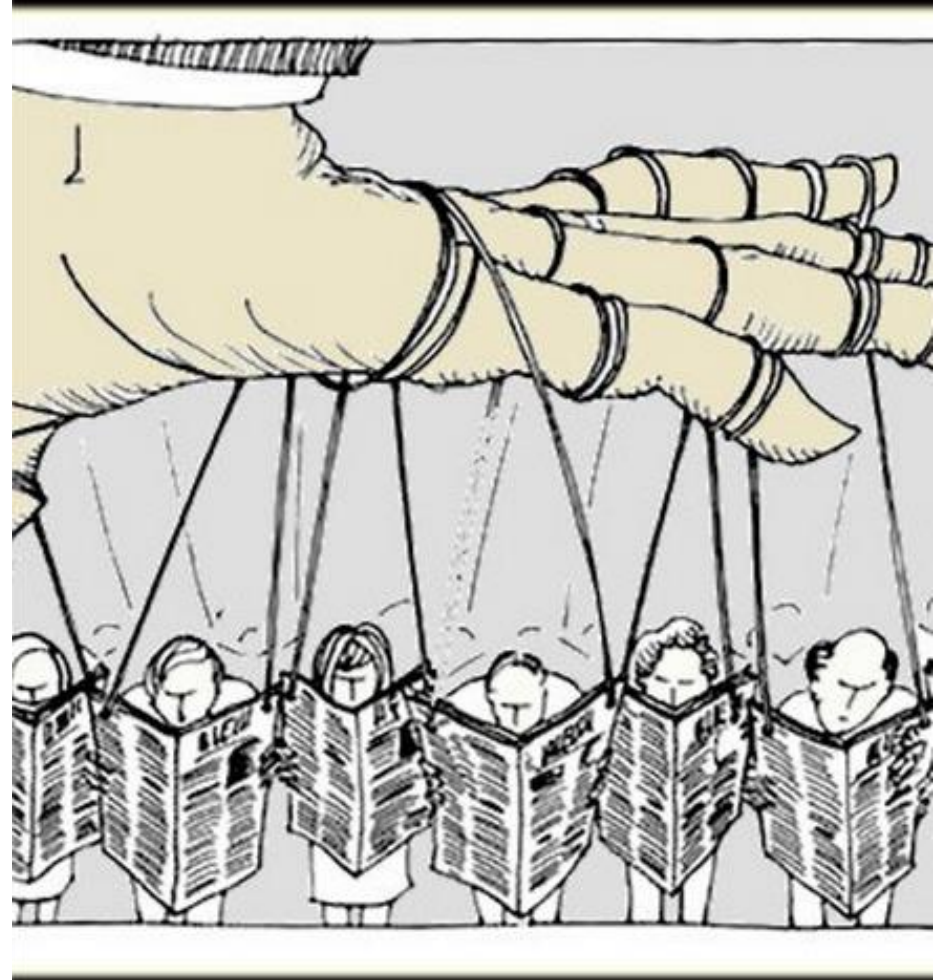
**Objectives:**

❖ XAI requirements related to AI-driven misinformation detection

❖ XAI by design using Knowledge Graphs

❖ XAI models for detecting information manipulation

❖ Generate creative and engaging explainability visualisations

❖ Personalize XAI to end-users skills and topic affinity

# Perceptions and Explanations

chist-era

## Politifact

**Facebook posts**
stated on July 8, 2020 in on social media:

Only 56 coronavirus deaths a day was enough to collapse the world economy, even though thousands of people die daily from tuberculosis, hepatitis B, and other diseases.

PANTS ON FIRE!
POLITIFACT TRUTH-O-METER

**Viral image**
stated on June 17, 2020 in a Facebook post:

"Four kids who took the coronavirus vaccine died immediately."

FALSE
POLITIFACT TRUTH-O-METER

PUBLIC HEALTH | FACEBOOK FACT-CHECKS
CORONAVIRUS | VIRAL IMAGE

## Fact-checking

**Facebook posts**
stated on June 24, 2020 in an image:

Google and Apple "put a covid tracker on every phone."

HALF TRUE
POLITIFACT TRUTH-O-METER

TECHNOLOGY | FACEBOOK FACT-CHECKS
CORONAVIRUS | FACEBOOK POSTS

No #SocialDistancing and it can even transmit from one to another. *See the video.*
scam

WHO says coronavirus patients don't need to be isolated or quarantined?

FALSE

## Snopes

### Will COVID-19 No Longer Be Classified as an Epidemic?

Report citing the CDC claimed coronavirus death rates in the U.S. were plummeting.

By Nur Ibrahim
Published 14 July 2020

## Fullfact

HEALTH / CORONAVIRUS
### This letter claiming to be from a doctor in Surrey is full of errors

23 JULY 2020

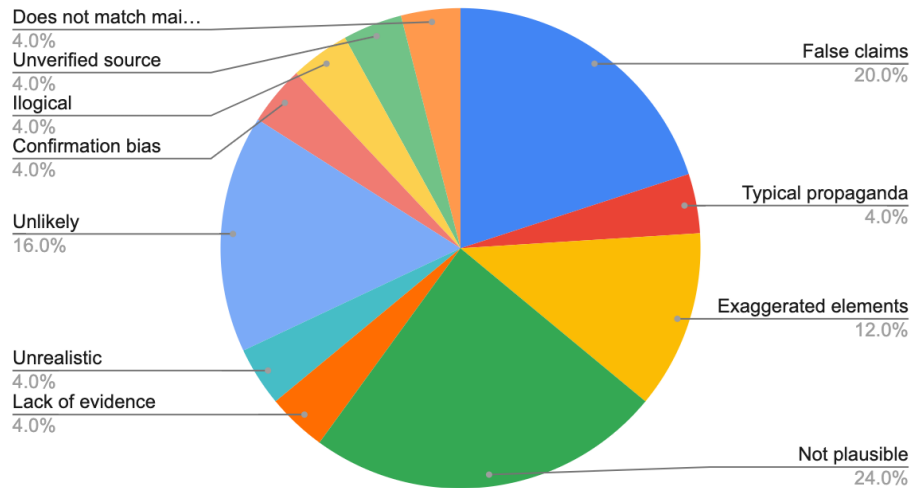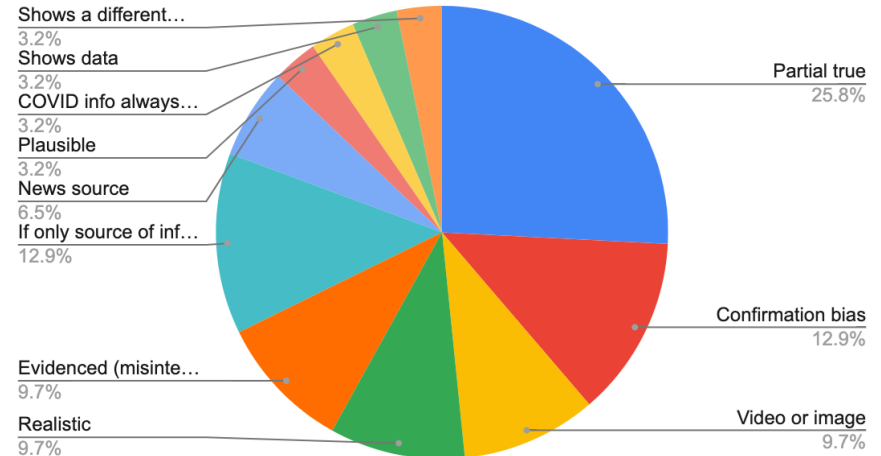| WHAT WAS CLAIMED | OUR VERDICT |
|---|---|
| A letter from a consultant at a hospital in Surrey claims that hospitals are empty and that there is no pandemic. | There are numerous inconsistencies in the letter which suggest it is not from someone working at a hospital in Surrey. |

## https://coronacheck.eurecom.fr/

- ❖ **Content type: letter, opinion**
- ❖ **Endorsement (accountable organisation)**
- ❖ **False claims**
- ❖ **Misleading strategy**
- ❖ **Fallacy detected**

**Reasons to believe**

- Shows a different… 3.2%
- Shows data 3.2%
- COVID info always… 3.2%
- Plausible 3.2%
- News source 6.5%
- If only source of inf… 12.9%
- Evidenced (misinte… 9.7%
- Realistic 9.7%
- Partial true 25.8%
- Confirmation bias 12.9%
- Video or image 9.7%

**Reasons to NOT believe**

- Does not match mai… 4.0%
- Unverified source 4.0%
- Ilogical 4.0%
- Confirmation bias 4.0%
- Unlikely 16.0%
- Unrealistic 4.0%
- Lack of evidence 4.0%
- False claims 20.0%
- Typical propaganda 4.0%
- Exaggerated elements 12.0%
- Not plausible 24.0%

# Knowledge Graph Centred Approach

- ❖ **KG of news items - ASRAEL**
- ❖ **KG of claims (and fact-checked claims) – ClaimsKG**
- ❖ **Domain-specific KG**
  - ✓ Covid-19
  - ✓ Climate Change
- ❖ **Vocabulary of propaganda techniques**
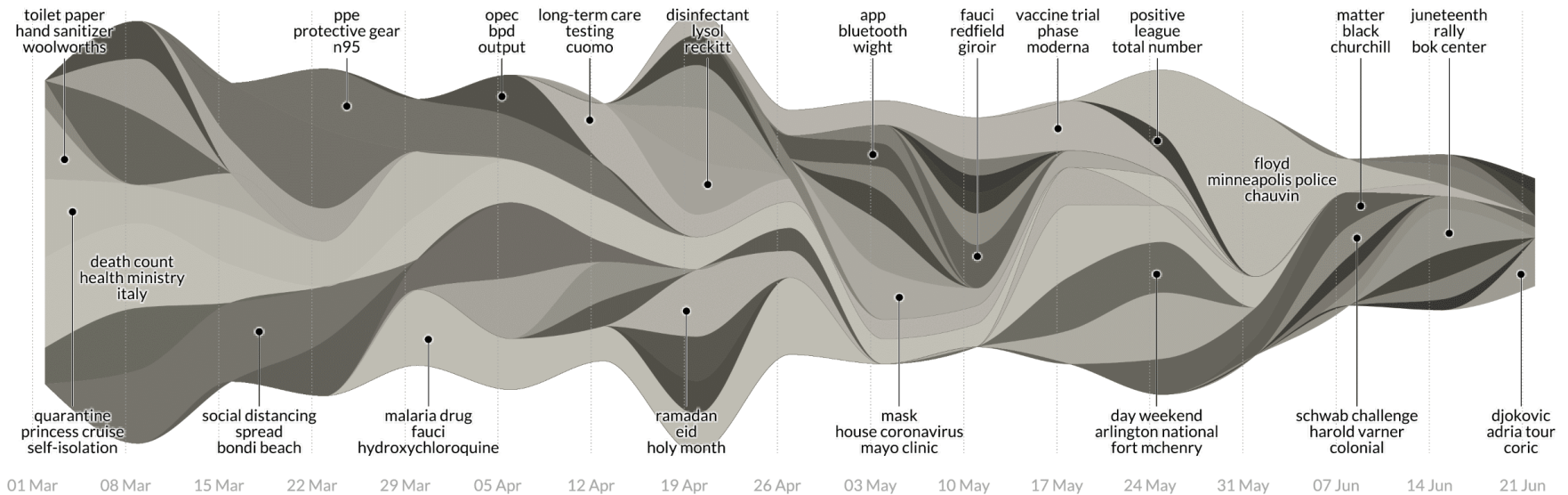- ❖ **Vocabulary for sentiment, emotions, bias**
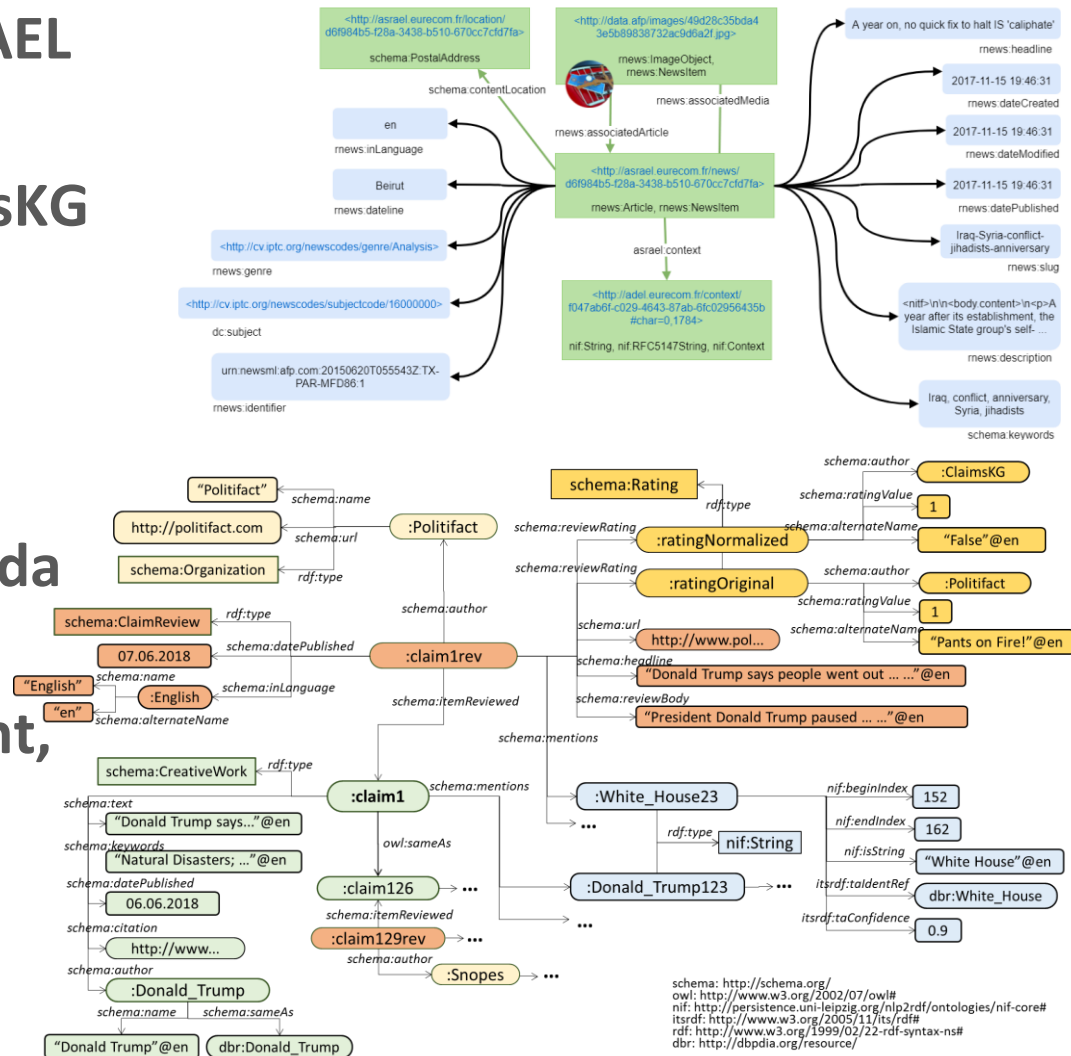
**https://www.weblyzard.com/corona-mood-barometer/**

The Corona Mood Barometer — Emotions and Associated Topics in May 2020
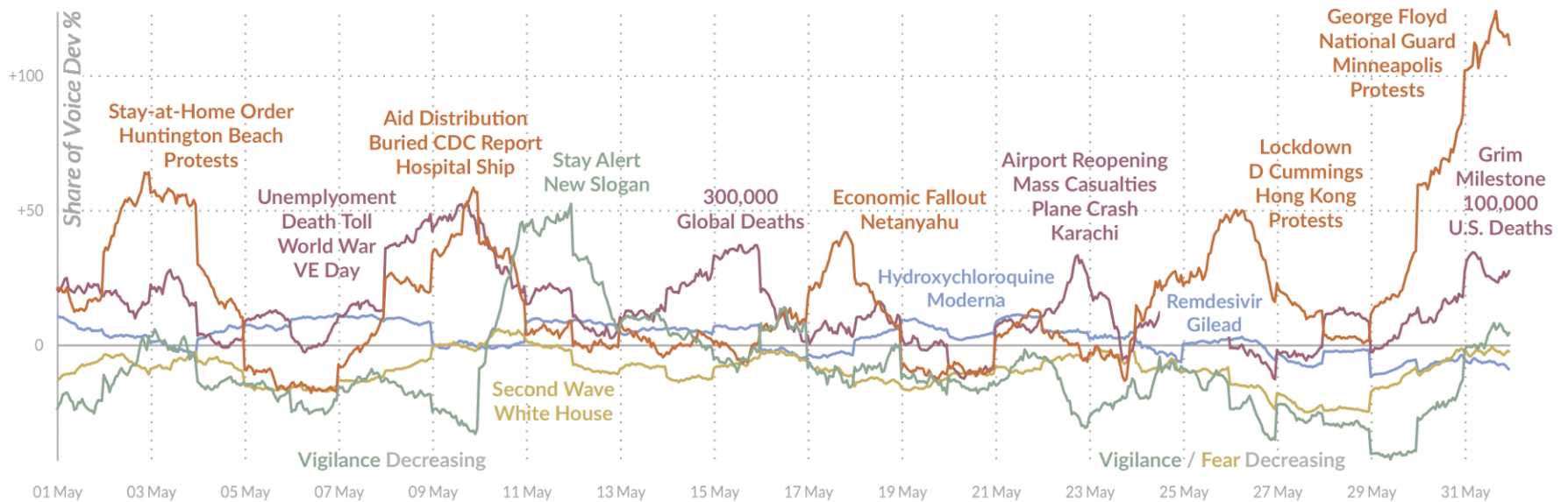
**https://www.weblyzard.com/corona-mood-barometer/**

**OPINIONS, INTENTIONS, FREEDOM OF EXPRESSION, ... , AND OTHER HUMAN ASPECTS OF MISINFORMATION ONLINE**

CHI 2021 - May 7th

CHI 2021

**https://events.kmi.open.ac.uk/misinformation/**

## ABOUT

Tools and online services are continuously being developed to support social media users, journalists, and policymakers to limit the harm caused by misinformation.

As we have experienced, the expected impact of online solutions is quite often hampered by limitations associated with lack of explainability, complex user interface, limited datasets, restricted accessibility, biased algorithms, among others issues that can confuse, overwhelm, or mislead users in their own ways. These limitations lead to **ethical implications that are typically neglected when new digital solutions to tackle misinformation are conceived**. As a wicked problem, limiting the harm caused by misinformation requires merging multiple perspectives to the design of digital interventions, including an understanding of human behaviour and motivations in judging and promoting false information, as well as strategies to detect and stop its propagation without unduly infringing on rights or freedoms of expression.

This hands-on workshop proposes to **unpack the state-of-the-art on social, societal and political studies, as well as socio-technical solutions to stop misinformation**, challenging the participants to

9

# Role of the CHIST-ERA Support

**Interdisciplinary nature of the research**
- ❖ Knowledge graphs for XAI
- ❖ Misinformation
- ❖ Story evolution and visualization
- ❖ Psychology of explanations
- ❖ Computational creativity
- ❖ User-based co-creation and evaluation

**Responsible research and innovation**
- ❖ Open science, open source development, sharing datasets
- ❖ Numerous scientific challenges are organized

# Questions ?