



18th May
2022
9:00 - 17:30
CEST

©metamorworks - Getty Images

AI: MITIGATING BIAS & DISINFORMATION

Theme Development Workshop

Identify common goals between academia and industry as well as other relevant stakeholders, and define promising approaches for European research and innovation in Trustworthy AI.

Organising Committee



Workshop programme

Due to Covid 19, the workshop will be held online with a mixed programme of presentations and in-depth discussions about specific sub-topics in smaller groups (Breakout sessions). This gives you the opportunity to discuss with selected experts and contribute to the strategic research and innovation agenda for AI in Europe.

- 09:00 – 09:15 Welcome & Objectives
- 09:15 – 09:30 Inoculation against misinformation
Prof. Dr. Sander van der Linden
University of Cambridge
- 09:30 – 09:45 Responsible AI approaches to address bias and misinformation
Prof. Dr. Virginia Dignum
Umeå University
- 09:45 – 10:00 Coffee Break & Socialising
- 10:00 – 11:30 Parallel Breakout sessions
- 11:30 – 12:30 Plenary presentation of key findings from the Breakout sessions
- 12:30 – 13:30 Lunch break & Socialising
- 13:30 – 13:45 Democracy and Knowledge in the Digital Age
Miguel Poiares Maduro
Chair of the European Digital Media Observatory (EDMO)
- 13:45 – 14:00 The challenge of finding the keyboard terrorist
Mijke van den Hurk
Police of The Netherlands & Utrecht University
- 14:00 – 15:30 Parallel Breakout sessions
- 15:30 – 15:45 Coffee Break & Socialising
- 15:45 – 16:45 Plenary presentation of key findings from the Breakout sessions
- 16:45 – 17:30 Closing & Socialising

Please register here.

We invite the community to suggest further topics of interest for the breakout sessions. Please use the [online application form](#) for your suggestions.

Breakout sessions

Breakout session 1:
The "arms race" nature of DeepFake detection
This session will focus on the issue of new generation DeepFake detection models that can evade their detection, leading to an "arms race" of AI methods and possible ways to stop or at least slow down this arms race by regulatory measures.

Breakout session 2:
Explainability aspects in AI for disinformation
The session will focus on identifying potential effective strategies and new approaches to fill the gap between existing AI explanation capabilities offered by current systems and AI explanations that are needed in the media domain.

Breakout session 3:
Science Communication with and on AI
This session will address how transfer of scientific insights to society has changed in digital media environments in recent years. Specifically, the role of AI technology in science journalism will be discussed, complemented by strategies for specifically increasing trust in AI for society.

Breakout session 4:
A social cognitive perspective to AI and misinformation
This session will focus on the challenges as well as opportunities to better integrate a social and cognitive perspective to the analysis of misinformation spread and social dynamics in human-AI hybrid systems.

Breakout session 5:
Abusive Language Detection and Comment Moderation
This session will investigate the potentials and risks of automated content moderation with a specific focus on abusive language detection. Besides algorithmic challenges, this session will also address ethical and legal aspects.

Breakout session 6:
Automation in Online Media
This session will deal with technical, societal, regulative, and ethical challenges modern AI-based automation technologies pose to communication in social networks with respect to disinformation and manipulation.

Breakout session 7:
Measuring Polarisation, Radicalization, and the emergence of Echo Chambers in online debates
This session will discuss how to identify, measure and characterise the interplay between users' social interactions in social media and the content they share and consume, also in terms of polarisation and radicalisation.

Breakout session 8:
Dataset sharing and governance in AI for disinformation
This session will focus on the balance between ethical, regulatory and technical aspects of disinformation-related dataset sharing. These aspects need to be handled jointly to publish datasets that are technically useful and socially acceptable to share.

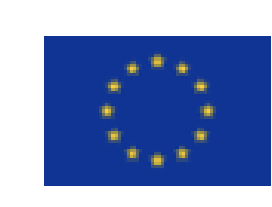
Breakout session 9:
What is bias and when is it bad?
This session will focus on how bias can be beneficial. It is often assumed that biases are bad, but they are relative to some criteria that can change over time or context. These criteria will be looked at in more detail in the session, e.g., regarding gender bias.

Breakout session 10:
Information vs. Disinformation?
This session will focus on the concepts of "unformation" vs. disinformation. False information leads to disinformation, but information can be far more misleading if it does not contain the whole story. This "unformation" gives an incomplete view of reality and leads to misconceptions, for example in war propaganda.

Breakout session 11:
SafetyTech
This session focuses on SafetyTech, technology that protects individuals in digital spaces. To avoid consequential problems arising from techno solutionism, we must find solutions that truly enrich the online experience, while respecting our values and freedoms.

Breakout session 12:
Online manipulation
This session explores how human decision-making in highly mediated digital environments becomes the target of actors that are inclined to abuse their power at the expense of the very fabric of society.

©Tony Studio - Getty Images



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under GA No. 951511 (AI4media), 825619 (HumaneAI-Net), 952215 (TAILOR) and 952070 (VISION).