

# ADVANCED RESEARCH WORKSHOP

Advanced Research Workshops (ARWs) contribute to the assessment of existing knowledge on important topics in security-related civil science and technology and serve to build networks among scientists from NATO and its Partner countries

Organized by  
Herbert Simon Society and KSE



Rome, November 16-17, 2021

**Fighting disinformation in a pandemic world:  
*the role of AI and cognitive sciences***



*This activity  
is supported by:*

The NATO Science for Peace  
and Security Programme

## Book of Abstracts

### Day 1

#### Introduction

**RICCARDO VIALE** (Herbert Simon Society and University of Milano-Bicocca)  
*“Fake AI to Fight Fake News?”*

Initial Simulative AI failed to be successful for an unrealistic model of mind (mind as a Turing Machine) and of environment (characterized by expected probability). Embodied Cognition and Ecological Rationality (New Cognitivism) offers an alternative realistic psychological model for AI. Current AI that is Machine or Deep Learning (ML/DL) have no simulative aim and its goal is only computing power. ML/DL may recognize only inductive patterns of past data about stable events but not about unstable events. Some ML/DL algorithms pretended to predict the future but they resulted to be serious failures as in the case of Google Flu Trends and IBM's Watson. Since disinformation sites are varying rapidly ML/DL may help to identify only few stable fake news accounts but not most of unstable disinformation behaviour. New Cognitivism may help to fight disinformation and fake news by boosting the users. This may be achieved changing the news feed algorithms that are responsible of epistemic bubble and polarization and introducing new digital architectures of choice that nudge critical approach to news. Finally Self-Nudges that include digital literature, risk literature, debiasing techniques and proper implementation intentions may boost user to approach critically fake news.

#### SESSION 1

*Disinformation and the human mind: socio-cognitive remedies against fake news*

**GERD GIGERENZER** (Max Planck Institute for Human Development and Director of Harding Center for Risk Literacy, University of Potsdam) - **keynote speech**  
*“Fighting Disinformation With Risk Literacy”*

Disinformation means sharing false content with the intent to mislead. Each new revolution in communication technology, from the printing press to the internet, has been followed by a wave of disinformation. Governments can respond to the problem in three ways: restricting people's choices by regulation (paternalism), steering people's choices (nudging), or increasing risk literacy so that people can make informed choices themselves (boosting). Although some degree of regulation is necessary, boosting is the most effective and sustainable measure against disinformation. As the Covid-19 pandemic has shown, many people are confused by the daily numbers they read, contributing to the rise of conspiracy theories and mistrust of science. Risk-literate citizens who understand statistics, uncertainty, and the importance of science would be more immune to this disinformation.

*Literature:*

Gigerenzer, G. (2014). Risk savvy: how to make good decisions. Penguin Press.

Gigerenzer, G. (2022). How to stay smart in a smart world. Penguin Press.

**PETE LUNN** (ESRI- The Economic and Social Research Institute, Dublin)

*“Behavioural Public Policy and Misinformation”*

Increasing concern about misinformation is often expressed as the problem that people are inclined to believe untrue claims spread on social media. I argue that people have always believed untrue claims told to them by others and that, therefore, this does not accurately identify the problem. Rather, the change brought by social media is that powerful commercial and political influences can now access everyday social interactions, which used to take place privately. These influences can record, analyse and interfere in such interactions, without being seen. Coping with this is the modern policy challenge.

**HUGO MERCIER** (Institut Jean Nicod, Paris)

*“Why does fake news have so little impact?”*

Fake news is often described as a major threat. In fact, it represents a tiny share of our information diets, it is mostly shared by people who already have strong political opinions, and it is unlikely to change anyone’s mind. Interventions that aim at fighting fake news are bound to have vanishingly small effects compared to interventions that aim at improving trust in reliable sources of information. Fake news is only worrying as a symptom of deeper problems, not as their cause.

**FABIO PAGLIERI** (ISTC-CNR, Rome)

*“The war on knowledge:critical thinking education as a strategic asset in the fight against disinformation”*

I will begin by arguing that disinformation is best understood as an alarming symptom of more fundamental problems in our media ecology, rather than as their root cause. This will prompt me to broaden the scope of the phenomena that are relevant to understand and curtail disinformation dynamics, in two complementary directions: at the individual level and at the infrastructural level. Concerning the individual predicament of online users, I will discuss the increasing phenomenon of social media fatigue, its relationship with context collapse, and the reasons behind a widespread tendency to engage in hate listening, i.e. deliberate exposure to problematic information flows that could have been easily avoided. With respect to the technological infrastructure of online interactions, I will briefly discuss the complex interplay among data colonialism, attention economy, the unfair distribution of dividends in user-generated contents, and the resulting need for private tech companies to promote Internet addiction in users. All of this is essential to clarify exactly what kind of critical thinking education is needed, to empower users to address these problems. I will offer three notable examples of successful interventions, based on learning by doing (inoculation against misinformation via online games), learning how to behave in public online (the Italian Manifesto of Non-Hostile Communication), and leveraging virality to promote reliable information (the Taiwanese "humour over rumor" strategy to fight Covid-related disinformation). In spite of these success stories, my analysis will end up on a sobering note: while appropriately designed educational interventions are and will remain necessary to fight disinformation, they are not sufficient, unless paired with more (and better) regulations of the digital media sector.

## SESSION 2

### *AI against disinformation: open problems and viable solutions*

**SANDER VAN DER LINDEN** (University of Cambridge, UK and Director of the Cambridge Social Decision-Making Lab) - **keynote speech**  
*“Inoculating Against Fake News”*

Much like a viral contagion, false information can spread rapidly from one individual to another. Inoculation theory therefore offers a natural basis for developing a psychological “vaccine” against misinformation. In a series of randomized lab and field studies, we show that it is possible to pre-emptively “immunize” people against disinformation. I’ll showcase several award-winning interventions we empirically evaluated—with governments and social media companies—to help citizens recognize and resist unwanted attempts to influence and mislead.

**MARIAROSARIA TADDEO** (University of Oxford and Programme Director of the DPhil in Information, Communication and the Social Sciences, Oxford Internet Institute)  
*“Cybersecurity and the Ethics of Information”*

Defence agencies across the globe identify artificial intelligence (AI) as a key technology to maintain an edge over adversaries. As a result, efforts to develop or acquire AI capabilities for defence are growing on a global scale. Unfortunately, they remain unmatched by efforts to define ethical frameworks to guide the use of AI in the defence domain. This article provides one such framework. It identifies five principles -- justified and overridable uses; just and transparent systems and processes; human moral responsibility; meaningful human control; reliable AI systems – and related recommendations to foster ethically sound uses of AI for national defence purposes.

**IRYNA GUREVYCH** (Technical University of Darmstadt, Germany and Head of Ubiquitous Knowledge Processing-UKP Lab)  
*“Natural Language Processing for AI-assisted Fact-Checking”*

Human fact-checkers cannot cope with the massive amount of misinformation spread and amplified via social media. In this talk, we highlight how our natural language processing (NLP) techniques can support fact-checking and discuss how to collect the data needed for such systems. We demonstrate the power of our approach, based in evidence retrieval and stance detection, on the FEVER shared task. We then explore the shortcomings of the previous era of NLP-based fact-checking and how modern AI systems can combat the ongoing infodemic. By creating AI systems that consider real world-claims, reasoning gaps in misinformation, fallacious reasoning, and multiple types of data, we can use technology to protect ourselves against fake news in its many forms.

**ELENA MUSI** (University of Liverpool)

*“Human Computer Interaction to counter fake news”*

This talk focuses on human computer interaction tools as means to teach critical thinking for data literacy. First, I will summarize the results of the UKRI funded project Being Alone Together: Developing Fake News Immunity where we propose to counter misinformation providing citizens with the means to recognize misinformation leveraging fallacy theory taught by two chatbots. Second, I will report on an ongoing project aimed at building a classification framework for human computer interaction tools to teach critical thinking for data literacy.

## Day 2

### SESSION 3

*Disinformation as a sociotechnical challenge: vulnerabilities and opportunities*

**WALTER QUATTROCIOCCI** (Sapienza University of Rome and Director of Center of Data Science and Complexity for Society, Sapienza University of Rome)-**keynote speech**  
*“Social Dynamics on Social Media”*

Do echo chambers exist on social media? By focusing on how both Italian and US Facebook users relate to two distinct narratives (involving conspiracy theories and science), we offer quantitative evidence that they do. The explanation involves users' tendency to promote their favored stories and hence to form polarized groups. Confirmation bias helps to account for users' decisions about whether to spread content, thus creating informational cascades within identifiable communities. At the same time, aggregation of favored information within those communities reinforces selective exposure and group polarization. We provide empirical evidence that users tend to assimilate only confirming claims and ignore apparent refutations because they focus on their preferred narratives.

The COVID-19 pandemic was the perfect storm for this phenomenon and the WHO coined the term infodemics to refer to the overabundance of information. We explored these processes during the pandemics' initial phase finding that reliable and questionable information spread similarly. We conclude the presentation by showing how different social media platforms (Facebook, Twitter, Gab, and Reddit) elicit very different polarization dynamics.

**ALEKSANDRA CISLAK** (SWPS University of Social Sciences and Humanities-Warsaw)  
*“Motivated group identities and misinformation acceptance”*

As the informational deficit model no longer provides a good exploratory framework of the anti-science attitudes, there is a growing need to seek and verify group-based motives behind misinformation acceptance. We probe the motivated group identities, which are focused on the in-group’s image enhancement.

We review extensive empirical evidence showing that motivated social identities are linked to:

- (1) vaccine, COVID-19, and climate change conspiracy theories.
- (2) mistrust toward scientists, climate skepticism, and vaccine avoidance.
- (3) support for the public policies that may undermine national position (e.g., Brexit/Polexit, anti-environmental or solutions undermining public health).

Thus, we argue that tackling misinformation acceptance requires tackling underlying group-based motives (not only information and knowledge provision).

**ALEKSANDRA CICHOCKA** (University of Kent, UK)  
*“Conspiracy theories: beliefs and dissemination”*

Conspiracy theories are attempts to explain causes of significant social and political events with claims of secret plots by powerful actors. There is increasing evidence that they pose a threat to democracy and social cohesion. Thus, we need to understand factors that make people susceptible to believe in and share conspiracy theories. I reviewed evidence for conspiracy beliefs being associated with epistemic (understanding one’s environment), existential (being safe and in control of one’s environment), and social (maintaining a positive image of the self and one’s social groups) psychological motives (Douglas, Sutton, & Cichocka, 2017). I also discussed emerging research on social motives associated with sharing conspiracy theories on-line.

**PIETRO GRAVINO** (Sony CSL Paris)  
*“Towards Post-Fake Era”*

Disinformation is as old as lies, and it has been used as a weapon in different shapes, evolving as Information Technologies did throughout the ages. Despite the dramatic improvement of information availability, it can have a dangerous impact on our societies, even when it involves only relatively small minorities. Democracies are currently struggling to find a way to deal with the problem without hindering the defining values of democracy itself.

But while we cannot improve information quality without entering the battlefield of opinions, we can do a lot to enhance information accessibility. We can redesign Information Technologies to make social dialogue more transparent, understandable, and healthy. The "Infosphere" research line at Sony Computer Science Labs is tackling these challenges through different initiatives: (i) the detection of unmet news demand that might trigger disinformation production; (ii) the building of bridges between polarised factions through new recommender systems and transparent, shared-value reputation system for news outlets; (iii) the visualization of the social

dialogue to improve citizens awareness about the different points of view; (iv) the study of "divisive news" instead of "fake news", whose definition can always be questioned. Our final aim is to improve our societies' information dynamics through new IT tools shaped around human information processing, both at the individual and at the collective level, to contrast those features that are resulting in dangers for our democracies.

## **SESSION 4**

### *Media policy and disinformation: political and ethical aspects*

**STEPHAN LEWANDOWSKY** (University of Bristol) - **keynote speech**

*“COVID-19 vaccine hesitancy: origins and possible responses”*

The COVID-19 pandemic has been accompanied by increasing trust in science and scientists by the public in many European countries and vaccine uptake in most countries is increasing and hesitancy declining. However, a sizable share of the public remains hesitant and the opposition to vaccines among those people is hardening. One important driver of hesitancy is political worldview, with opposition on the political right being particularly prominent. To overcome hesitancy, it is important to use worldview-congruent messengers to engage hesitant individuals. In addition, the public can be protected against misinformation through cognitive inoculation. Inoculation provides people with the skills to recognize questionable information.

**DAMIAN TAMBINI** (The London Schools of Economics and Political Science)

*“Algorithmic Pluralism: censorship in Information Warfare”*

Recent cases of disinformation show that information warfare is asymmetric: democracies are more vulnerable than authoritarian countries to disinformation campaigns because public opinion and deliberation enable democratic states to act. Information warfare and disinformation through social media have spurred new forms of social media regulation which may undermine the free flow of expression and opinion. Following allegations of foreign disinformation campaigns in Europe for example online safety legislation has been drafted in the UK and the Digital Services Act published in the EU. This talk examines the background to the new approaches to speech regulation in Europe and argues that long term resilience of democracies should be based on a new approach to plurality of media algorithms in the age of AI, not censorship as a defensive posture.

**SARA RUBINELLI** (*University of Lucerne*)

*“Building resilience against mis/disinformation: WHO competency framework for infodemic management”*

The presentation focuses on the WHO framework for infodemic management (IM). It identifies competencies ranging from monitoring and measuring infodemics to designing, conducting, and evaluating appropriate interventions as well as building resilience to infodemics among populations. The competency framework contains primary domains, each of which is comprised of main activities, related tasks, knowledge and skills. Its purpose is to assist institutions in strengthening their IM capacities and implementing effective IM processes and actions according to their individual contexts and resources.

**ANNI STERNISKO** (*New York University*)

*“The social complexities of conspiracy theories”*

To combat the spread of conspiracy theories, we must understand the different psychological pathways that lead people to believe and disseminate them. In the present talk, I discuss a motivational framework that captures the complexity of conspiracy theories and allows us to predict when and what kind of conspiracy theories gain momentum in society. I then present studies that applied this framework to understand the spread of conspiracy theories in different contexts. For instance, I discuss my recent work on the spread of conspiracy theories related to COVID-19 and on psycholinguistic markers of conspiracy theories in online social networks. Last, I discuss opportunities for future research.