EU DISINFOLAB
2022 CONFERENCE DOSSIER

EU DISINFO LAB

EU DisinfoLab 2022 Conference Dossier

This Dossier summarises some of the sessions held during the EU DisinfoLab 2022 Annual Conference, which took place on 25 and 26 October 2022. Over the two days, dozens of experts from the counter-disinformation community shared their expertise on stage with hundreds of participants, and this document – together with the recordings of the video sessions available online – come as a result.

This Dossier aims to recap the most salient moments of the conference to build a legacy of what has been discussed, exchanged, and learned. The next pages merely reflect the position of the speakers on the various topics that, given the nature of the conference, are not necessarily exhaustive or equate to an endorsement.

This Dossier is authored by conference rapporteurs Melanie Katharina Döring and Iana Pancenco.

TABLE OF CONTENTS

KEYNOTE – VACCINATION AND DISINFORMATION

*Opening keynote by **Dr. Raed Arafat**, Secretary of State, Head of the Department for Emergency Situations, Ministry of Internal Affairs, Romania*

You can watch the recording of this session here.

At the EU DisinfoLab 2022 Annual Conference, Dr. Arafat, Secretary of State and Head of the Department for Emergency Situations at the Interior Ministry in Romania, reported on the extensive disinformation campaign against his person amid the COVID-19 pandemic. Anti-vax activists and a local TV station with some of the highest ratings in Romania, attacked him following the adoption of containment measures.

The backlash against coronavirus-related protective initiatives was massive, and Dr. Arafat was turned into a scapegoat by influencers, government opponents, and the media. These actors tried to influence medical decisions, which hugely impacted the population. National broadcasters regularly hosted obscure opinion leaders like fashion designers and priests on primetime talk shows, offering a platform to these actors. Other media outlets further amplified their unscientific claims, which contributed to legitimising and granting them credibility.

> *"Not social media but claims taken to parts of the regular media were the problem: inviting those influencers and attacking certain people (...). The problem is the discussion with anti-vaxxers,"* said Dr. Arafat.

Dr. Arafat called out the correlation between the fact that these TV shows were the highest-rated among senior citizens during the pandemic and that the fourth COVID-19 wave was the deadliest in Romania. In fact, despite the vaccine's availability, only 24% of elderly people were vaccinated.

The health expert's attempts to dispel rumours and accusations with scientific evidence, fact-checked content, and discussions were rarely effective. The constant stream of fabricated stories and harassment silenced many of Arafat's colleagues, who stopped defending their positions in view of the death threats and abuse they were met with.

While the Secretary of State won court cases against the TV company responsible for the reported disinformation campaign, the small sum of money he received did not outweigh the damage done, especially as these malign and complicit actors continue to earn greatly in advertising revenues.

KEYNOTE – ANTI-ABORTION DISINFORMATION: COORDINATED OLIGARCHS AND RELIGIOUS GROUPS

*Opening keynote by **Neil Datta**, Executive Director, European Parliament Forum for Sexual and Reproductive Right*

You can watch the recording of this session here.

Neil Datta, Executive Director of the European Parliament Forum for Sexual and Reproductive Rights, discussed how anti-abortion movements counter-balance the liberalisation in reproductive rights.

This trend started systematically in 2010 when the level of contestation and anti-abortion activism increased across Europe. Then, a new kind of organisation emerged that knows how to access power, draft legislation, and engage in litigation. They follow a targeted harassment strategy against pro-abortion activists and organisers. On this front, court action – rather than political acts – is key, as the overturning of Roe v. Wade in the United States or the 2020 regression on abortion in Poland demonstrated.

"This is a war on gender equality," said Neil Datta.

These players are well-connected transnationally, maintain ties with peers abroad, and wait for the right time to run campaigns in their countries. To further legitimise their view, they built a whole disinformation universe filled with pseudo-science, fake university degrees, compliant media, and alleged fact-checkers. Religion is just a pretext for anti-abortion activists, granting credibility to a modern political power project that intends to gain leverage among religious people through various means that range from targeted astroturfing to petitioning via the ultra-conservative platform CitizenGo.

The new extreme and very vocal movement uses gender to sanitise anti-human rights positions and is not grounded in liberal democracy or the rule of law as gender becomes part of the national narrative. These stakeholders' bond with ultra-nationalists or ultra-capitalists that oppose all state intervention. Funding for these researched organisations roughly quadrupled to 700 million USD from 2009 to 2018, stemming from the U.S., Russian oligarchs, and highly networked Europeans.

Last but not least, anti-abortion activism is also part of the authoritarianism toolkit. For instance, Mr. Datta mentioned that the recently elected Italian government led by Giorgia Meloni has already prepared a draft law asking to grant foetuses legal status from conception within the Italian Civil Code. Although this trend is not new, there seems to be no solution, as Mr. Datta points out. On the contrary, these movements are expected to gear up along with an anti-gender equality agenda for the upcoming European elections.

PANEL – OSINT: LESSONS FROM UKRAINE

The widespread use and development of OSINT techniques in the context of the war have proven its potential for countering dis- and misinformation, and documenting war crimes, among others. What have we learnt and is it enough to bring accountability?

Chair: **Inês Narciso**, Researcher, MediaLab, ISCTE-IUL
Speakers: **Benjamin Strick**, Investigations Director, Centre for Information Resilience; **Elise Thomas**, Senior OSINT Analyst, Institute for Strategic Dialogue

You can watch the recording of this session here.

Open-source intelligence (OSINT) allows researchers to gather insights on what happens on the ground during ongoing war scenarios, such as in Ukraine, and makes it even possible to analyse war crimes as they are committed. However, the other aspect to consider when dealing with OSINT is the way in which bad actors can manipulate information. For example, COVID-19 conspiracy theorists are now spreading pro-Russian disinformation about the war – recycling narratives such as the secret biolabs theory.

Therefore, the fact that the same openly available information can be used for malign scopes shows the dark side of OSINT, e.g., geolocating ideological opponents. Ethical considerations must be made on what is shared and how. Moreover, transparency is a core part of OSINT. Once an investigation is finalised, one should be able to consult the sources and reproduce the analysis obtaining the same results.

"Let the facts tell the story, not the other way around," stated Benjamin Strick.

The rise of crowdsourcing contributes to the new trend of using OSINT for disinformation. Examples abound, from pro-Russian groups targeting LGBTIQ+ communities to people's addresses and personal information being tracked down and published. The circulation of these bits and pieces are the 'loudest' findings to reach the audience, compared to a real investigation with proper analysis, which takes time and resources.

On a more positive note, OSINT has a collaborative aspect that can help overcome other issues. A virtuous specimen is the Eyes on Russia map, which was built thanks to multiple organisations and is now used by judicial bodies to investigate war crimes.

MASTERCLASS – HOW NOT TO CONDUCT AN OPEN-SOURCE INVESTIGATION

OSINT tools are widely available, and many are eager to use them for good. But with great power comes great responsibility.

Speaker: **Aiganysh Aidarbekova**, Open-Source Investigator, Bellingcat

You can watch the recording of this session here.

OSINT researchers rely on sources available to everyone (e.g., ship or flight tracking websites to Google Maps, etc.). For this reason, such tools and methods can also be used by ideologically biased or extremist groups that misuse and abuse them. A sadly famous case is the Pizzagate conspiracy theories, in which believers showed the dangers of not conducting research inappropriately.

Those promoting disinformation will not rely on factual evidence but misinterpret facts and force them together to fill a predeterminate storyline. Examples of this reasoning abound among QAnon believers, where evidence is fabricated in search of causality where there is none.

> *"People make mistakes, but it is important to be honest and acknowledge when dis- and misinformation happens,"* said Ms. Aidarbekova.

Closed communities, the often-mentioned filter bubbles and echo chambers, reinforce the biased tendency to jump to conclusions. Therefore, people should search out of curiosity and be always open to asking questions. Lastly, it is key to make sure that findings and conclusions are accurate by challenging and comparing them.

PANEL – INFORMATION FRONTLINES

Recent investigations have shown how the Internet infrastructure is a new frontline to counter disinformation. Domain name abuses, registrars, servers, this infrastructure is central in accessing information. What do we know about it? Is there accountability in the field? Can we sanction and imagine future policies to prevent malicious actors from abusing it?

Chair: **Emma Le Mesurier**, Managing Director, Next Level Initiatives
Speakers: **Kevin Limonier**, Associate Professor in Slavic studies and geopolitics. Specialised in geopolitics of the Russian-speaking cyberspace, Université Paris VIII; **Alexandre Alaphilippe**, Executive Director, EU DisinfoLab

You can watch the recording of this session here.

People need to look at information space vulnerabilities the same way they look at infrastructure vulnerabilities (e.g., energy, security).

> *"We need to understand what is the cognitive critical infrastructure through which all of our information flows and, therefore, which has become and will continue to be the battleground,"* reminded Emma Le Mesurier.

Internet infrastructure consists of physical and immaterial infrastructure (i.e., internet protocols). Therefore, the internet is a set of standardisation protocols, a set of complex languages that carry data. Although internet protocols represent a geopolitical issue, the internet has been seen as something non-strategical, non-political, and deeply technical for a long time.

As data regularly goes through various routers until it reaches its final destination, the internet has a geopolitical shape. For instance, Russia has established a strategy for a territorial occupation that can be defined as a "digital occupation". The Kremlin re-routes data flows in order to control what goes in and out of the controlled territories. Thus, the internet's geopolitical shape can be manipulated.

According to Kevin Limonier, *"Internet protocols are a political issue, a geopolitical issue, and not only for carrying flux of data, but also for carrying information."*

Another attack on the internet infrastructure, mainly through the domain names infrastructure, is presented in EU DisinfoLab's Doppelganger investigation. The study reports on a large discovery of fake domains impersonating trustworthy news outlets, intentionally misspelling the domain names of well-known websites to redirect users to a different domain (e.g., from spiegel.de to spiegel.ltd). The lack of scrutiny or authority to shut down or monitor these domains is extremely concerning, in addition to the fact that new ones can be bought. Instead, the EU must keep the internet together and fight copyright and GDPR infringements. Massive disinformation campaigns are deceiving unaware citizens, and there is no easy way to take action against them on a grand scale.

PRESENTATIONS – DISINFORMATION ACROSS THE EU

Disinformation has many universal elements. It is also versatile and adapts to its surrounding context. What is the state of the art in different EU countries?

Chair: **Carlos Hernandez-Echevarria**, Head of Public Policy & Institutional Development, Maldita.es
Speakers: **Thanos Sitistas**, Senior Editor, Ellinika Hoaxes; **Jelena Berković**, Senior Advisor, Faktograf; **Hanna Linderstål**, CEO, Earhart Business Protection Agency and newly started Earhart Academy.

You can watch the recording of this session here.

The panel gathered fact-checkers and counter-disinformation experts, who shared insights based on their countries. In Greece, the quality of mainstream media deteriorated with the economic crisis, as less experienced journalists were hired, and facts were rarely verified. Today, conservative and far-right media thrive and enjoy popularity. Deceptive narratives focus on the pandemic and the war in Ukraine. Conspiracy theories promote prophecy-like claims that the Russians will give Istanbul back to the Greeks. Moreover, disinformation radicalises people and is sometimes amplified by media outlets, such as the claims that Ukrainian refugee women seduce men and break up families. The familiarity with Russia – with which Greeks share the same orthodox religious values – helps this propaganda spread and succeeds in the country, antagonising Western principles, according to Thanos Sitistas.

> *"Fact-checking itself is not the solution. Media literacy alone is not the solution. It's going to be a combination of all that,"* maintained Carlos Hernandez-Echevarria.

Even though fact-checking is a principle of good journalism, it has now become a new profession. As the area of Bosnia-Herzegovina, Serbia, Montenegro, and Croatia is linguistically similar; content easily travels across those countries. However, the Western Balkans are highly influenced by Russia, and the most common sources of disinformation are online media portals and social networks. While the DSA and the Code of Practice on Disinformation only apply to EU member states, the non-EU members in the Western Balkans should not remain behind. Otherwise, trust in institutions will be lost, and few critical voices will remain. People fall for disinformation because the natural root causes, such as building the capability to filter information, are not adequately addressed, mentioned Jelena Berković.

"Most of our decision makers are digital tourists, and they decide the future for digital natives, and that's really scary" warns Hanna Linderstål.

She mentions that, in Sweden, there needs to be more systemic understanding in the media.

Many journalists must understand how algorithms work or how influential social media is. Disinformation is rarely created in traditional media; the news is spun in a certain direction. From deceptions about money-winning schemes on WhatsApp to far-right messages originating from Russia ahead of the 2022 elections and disseminated on Snapchat, journalists should be educated on technology and how malign actors exploit it to easily find reachable communities of susceptible people. Important information needs to be simplified to ensure that complex issues are easy to grasp, as a big part of the population spends a lot of time online, and malicious people know how to target them.

MASTERCLASS – UNLOCKING THE POTENTIAL OF AI FOR VERIFICATION (VERA.AI)

Speaker: **Denis Teyssou**, Medialab R&D Editorial Manager, AFP

You can watch the recording of this session here.

Mr. Teyssou introduced vera.ai, an AI-assisted verification tool, explaining that the aim was to develop novel AI and network science-based methods that assist verification professionals throughout the complete content verification workflow. This includes text, images, audio, videos, and deepfakes.

> *"Our goal is to design new tools that are effective and can be used by thousands of people around the world;" said Denis Teyssou.*

With 14 partners, which include eight research partners and two commercial organisations, they expanded on the InVID & WeVerify plug-in, which currently has 80,000 weekly active users. Furthermore, they developed Truly Media, the disinformation monitoring backbone of the European Digital Media Observatory (EDMO), and the Database of Known Fakes (DBKF) to uncover reappearing fakes.

Among the modules is a biometric scanner to debunk deepfakes, mathematics-based forensic analysis, and link-sharing behaviour tools to identify communities that share disinformation, also using machine learning. Another effort was to create a tool that helps stakeholders solve concrete issues, such as finding better ways to present fact-checking results. For instance, GIFs are helpful to visualise doctored images before and after being manipulated, as in the case of fake pictures of Ukraine's First Lady Olena Zelenska.

Moreover, the goal of vera.ai is to hide technical complexity. Using AI means automating, simplifying interfaces, presenting evidence, and improving the overall fact-checking user experience. For this to work out, errors must be minimised, and there needs to be more awareness of these tools. Lastly, there are still many unsolved challenges – especially deepfakes and link-sharing behaviour – as well as limits when it comes to platforms and languages (with English remaining the main language).

PANEL – YOU SAY YOU'RE THE MEDIA, NOW ACT LIKE ONE!

From 'traditional media' and tabloids to citizen journalists and influencers, how can we ensure that all actors in our media system are responsible and accountable?

Chair: **Susan Morgan**, Independent Consultant
Speakers: **Luboš Kukliš**, Chief Executive, Slovak Council for Media Services, Chair, European Platform of Regulatory Authorities (EPRA); **Prof. Rasmus Kleis Nielsen**, Director, Reuters Institute for the Study of Journalism; Professor of Political Communication, University of Oxford; **Christophe Deloire**, Secretary General, Reporters without borders, and Chair, Forum on Information and Democracy

You can watch the recording of this session here.

The panel started by discussing the accountability of media actors, particularly social media platforms, in light of the Digital Services Act (DSA) and the European Media Freedom Act proposal (EMFA), and the role of and support needed for quality media in the information ecosystem.

The panellists explained that in the absence of set definitions, activists, governments, advertisers, brands, and digital companies often claim to be journalists for media freedom protection. However, there needs to be a distinction of whom this status should be granted to in order to sustain the future of journalism. Transparency, methods, and ethical compliance are key, as are self-assessments and external audits. While the EMFA could be a historic step forward, Article 6 of the proposal endangers editorial independence due to excessive influence by owners, advertisers, and other players.

Journalism can sometimes be part of the problem, yet it is certainly part of the solution. The media has a prime role in distributing reliable information. Yet, journalists can also get caught up in disinformation (e.g., from biased beliefs to actual interference from state actors and interest groups). Part of the media is also dubbed the "outrage industry", reporting non-objective and stereotypical content. Consequently, limiting content moderation to illegal content will lead to more false and misleading information in public space. Disinformation as such is not illegal, which makes it hard to regulate; competence and expertise are needed to detect and address it respectively. In this regard, the new European legal tools will be more efficient than those available at members state level.

Christophe Deloire defends that sanctions against propaganda media in Russia were a political decision. In the long term, it is up to independent regulatory bodies to make these decisions. However, disinformation must have consequences as there is an asymmetry between open democracies and despotic regimes that can launch manipulation offensives. Overall, new legal systems are needed to address this. Content regulation can function similarly to market regulation by having strong regulatory bodies implementing strong democratic safeguards. Public investments are required in order to address the challenges to democracy.

"It should not be the aim to silence people who lie or express hate (...). They have always existed. But in the past, they were on the side-lines, on the margins of the public space of society. [Now, they found ways to gain trustworthiness]. We have to get back to this old logic," continues Mr. Deloire.

Speaking of regulation, the new Code of Practice on Disinformation has been in effect since June and goes much deeper than the previous version, also regarding transparency, and sets KPIs. Implementation remains to be seen, as well as further development to enlarge its scope. While not part of the DSA, promoting good content is expressed in the Code of Practice. A greater focus on fostering good content rather than simply deleting bad content would also avoid pushing users to less-controlled platforms.

Lastly, it is positive that there is no media exemption in Article 17 of the DSA. If that were the case, a definition of media would be needed first, which is nearly impossible to get right. Such a rule threatens to make a comeback in the EMFA. To get procedural exemptions, what could work instead is trusted content creators, similar to trusted flaggers in the DSA. Further incentives – like public funding – could be offered if an entity shows proof of implementing journalistic standards.

GUEST INTERVENTION BY VĚRA JOUROVÁ, VICE-PRESIDENT OF THE EUROPEAN COMMISSION FOR VALUES AND TRANSPARENCY

You can watch the recording of this guest speech here.

> *"This war is not between Russia and Ukraine, but between dictatorship and democracy,"* said Ms. Jourová.

The world is experiencing the first digital war in history, and "we do not have a proper reaction," said Ms. Jourová. The pro-Kremlin disinformation machine promotes false information about missiles in Kyiv, while energy and food vulnerability remain constant topics. Disinformation actors manipulate to undermine free speech. Therefore, important actors should not only respond but also predict and tailor their communication accordingly.

> *"Autocrats hate active people. They hate demanding citizens. This is a missing chapter in our plan,"* continued Ms. Jourová.

The EU is supporting investments in research hubs. The European Democracy Action Plan offers a comprehensive approach that includes the role of media, society resilience, and legislative and non-legislative actions. Therefore, organisations like the EU DisinfoLab need to have access to data to be able to carry out their analytical work properly. But a certain degree of responsibility also regards the users who are or should be in control of what they consume.

Finally, quality media should be protected as it has the power to address and fight disinformation. The EU also needs proper legislation on transparency in political advertising to check unfair and opaque disinformation.

MASTERCLASS – CAN WE AGREE ON A DEFINITION OF DISINFORMATION?

The more everyone talks about 'disinformation,' the more clear and functional definitions matter in this space. How do different stakeholders and platforms define disinformation, and how does this affect their approach and our collective effort?

Chair: **Lutz Güllner**, Head of Division for Strategic Communications and Information Analysis, European External Action Service
Speaker: **Olga Belogolova**, Policy Lead, Influence Operations, Meta

You can watch the recording of this session here.

Disinformation often entails different meanings – from "fake news" to influence and information operations. As many different issues are included in the term, the key question that emerges is how we solve something if there is no agreement over its definition.

The phenomena themselves have been around for a long time. "Political warfare" is a recurrent term, and "active measures" have been used to reference the Soviet Union. "Propaganda" used to have a neutral definition – as in amplifying information – while the synonymous expression "public diplomacy" was explicitly used to distance the propaganda of the East from the West.

There are content-specific terms anchored in the concept of truth. However, determining what this is changes over time. The standard definition of disinformation is "false information deliberately created and intended to harm a person, group, or country". Misinformation is spread without the intent to deceive. Malinformation is "the spread of genuine information with the intent to harm, such as certain forms of leaks, harassment, and hate speech."

Content moderation solutions include removals, reducing reach, and informing the audience. Overt or covert campaigns – usually referred to as "information operations" or "influence operations" (IOs) – are where content-based solutions fail. After Russia created false persona campaigns, Facebook moved from these terms to Coordinated Inauthenticity Behaviour (CIB). These investigations adopt a behaviour-based approach to look for patterns of interactions, which do not require knowing the actor or classifying content as true or false. Success signals are if threat actors are forced into better operational security, adapting, or shifting to other platforms.

> *"The key to addressing disinformation is avoiding the pure and focusing only on content. Then it becomes a good vs. bad, black vs. white issue. A broadening with behavioural issues in combination with actor issues is the only way out of this political dilemma because you automatically infringe on the freedom of speech. We want to safeguard it with our tactics and procedures,"* said Lutz Güllner.

Deterrence in the IO space consists of discouraging malign actors by making their actions more costly, but it also includes disruptions like removing fake accounts, blocking domains, or declaring someone a persona non grata (PNG). Moreover, disclosure (i.e., sharing threat research with industry, governments, and civil

society), as well as transparency labels or notifications for users can be integrated, as well as soft actions (warnings) to bring violators back in compliance.

> *"We cannot solve the problem of disinformation, but we can build mitigating factors,"* continued Mr. Güllner.

PRESENTATIONS – HOW ARE OTHER PLATFORMS DEALING WITH DISINFORMATION?

The techniques and tactics of disinformation actors and the spread of misinformation vary in different online settings. How do different service providers perceive and mitigate these challenges?

Chair: **Isabelle Arnson**, Senior Policy Analyst, Tech Against Terrorism
Speakers: **Dimitar Dimitrov**, EU Policy Director, Wikimedia; **Bri Riggio**, Platform Policy Manager, Discord Inc.; **Laura Seritti**, Head of Public Policy Brussels, Snap

The session shed light on how three platforms deal with disinformation. For instance, the Wikipedia editing process involves patrollers, trusted users who get special editing rights. They also monitor certain articles in their respective language to notice changes made. However, the freedom of editing also opens doors to malicious people who aim to disinform, as in the infiltration of editing articles about Tunisian political regimes in French.

Discord envisions a three-tier moderation approach: user control, platform moderation, and content moderation. Therefore, misinformation manifests differently on Discord since it is a user-driven platform without advertisement. Therefore, there is no virality, no public feeds, and only private group-based interactions. However, the platform adopted a health misinformation policy after the spread of COVID-19 vaccine deceptions. This combined proactive and reactive methods: detection, third-party experts, and groups.

MASTERCLASS – GENDER-BASED DISINFORMATION: TACTICS, TECHNIQUES, AND PROCEDURES

Chair: **Maria Giovanna Sessa**, Senior Researcher, EU DisinfoLab
Speaker: **Lucina Di Meco**, Co-Founder, #ShePersisted

#ShePersisted is an initiative to fight gender-based disinformation against women in politics. Ms. Di Meco reports that 85% of women have witnessed or experienced abuse online, an issue we have known about for years. Ms. Di Meco has interviewed over 100 women in politics, female experts, and political activists, hearing horror stories about how they are being targeted online with hate and disinformative narratives.

#ShePersisted works on gendered disinformation as the spread of deceptive or inaccurate information or images against women political leaders, journalists, and female public figures that follow storylines that often draw on misogyny, as well as gender stereotypes around the role of women. This type of disinformation is designed to alter the public understanding of female political track records for immediate political gain. While gendered disinformation is rooted in misogyny, it goes well beyond, as authoritarians and illiberal actors weaponise negative stereotypes around women and strategically deploy them as a political weapon to eliminate political opponents and undermine democracy. Recurrent attacks question women's leadership abilities, sexuality, and character. Common disinformative narratives portray women as liars, unintelligent, traitors, too weak, too sexy, or not feminine enough.

Lucina Di Meco and her colleagues aimed at understanding the patterns, impact, and modus operandi of gendered disinformation campaigns against women in politics around the world. The case studies that they are about to publish explore how gendered disinformation has been used by political movements, and at times autocratic governments themselves, to undermine women's political participation, and to weaken democratic institutions and human rights. Crucially, the research also looks at the responsibilities and responses that both state actors and digital platforms have taken – or most often, failed to take – to address this issue.

> *"Social media has been failing women and democracy globally. (...) Digital platforms would love for us to think that this is just about the misogynist next door, or the sexism that has always happened (...). This is not what I'm talking about today,"* said Ms. Di Meco.

Digital literacy is insufficient to solve such a widespread, viral, and systemic problem. Some of the legal frameworks currently being negotiated are very promising. Freedom of speech is not in contradiction with stronger digital platform standards. In fact, they are a necessity as protection goes both ways. The status quo limits women's their freedom of expression online, and there needs to be awareness of this. Through their research, #ShePersisted saw how state-led gendered disinformation campaigns have been used to silence and deter women in politics from speaking out, stifling their calls for better governance. It is urgent to look at the business model of platforms that make profits out of hate and outrage and end up externalising the costs of the damage produced by their products on the entire society.

MASTERCLASS: ADDRESSING AD-FUNDED DISINFORMATION - HOW DOES AD TECH NEED TO CHANGE?

Speaker: **Clare Melford**, CEO, Global Disinformation Index

Disinformation often concerns the most emotionally charged content in order to drive engagement and generate ad revenue. A disincentive should be provided for those involved in online advertising to demonetise disinformation narratives — assessments of which sites are high risk for disinformation. This must be provided by neutral independent third parties with no stake in the current ad tech ecosystem.

The Global Disinformation Index produces the Dynamic Exclusion List (DEL) of global news publications rated high risk for disinformation. The DEL contains the worst offending websites and apps across multiple countries and languages and is continually updated to capture new disinformation sources and narratives. Ad tech companies and platforms can license GDI data to defund and downrank these worst offenders, thus disrupting the ad-funded disinformation business model. This list contains the highest density of disinformative, hateful, and overall bad sites that is then sent to ad companies.

> *"You have a right to free speech. You do not have a right to monetise everything,"* said Clare Melford.

Policies must target the monetisation of disinformation and disrupt the financial incentive for creating such harmful content. Ad tech publisher policies must be updated and enforced to address the breadth of disinformation narratives.

PRESENTATIONS – HOW TO MONITOR ALGORITHMS?

Is it possible to make an algorithm accountable? (We certainly hope so!) But how do we monitor and enforce this in practice?

Chair: **Mathias Vermeulen**, Director, AWO
Speakers: **Nelly Pailleux**, Chief Knowledge Officer, Check First; **Marc Faddoul**, Co-Director, Tracking Exposed; **John Albert**, Policy and Advocacy Manager, Algorithm Watch

You can watch the recording of this session here.

Recommendation algorithms are responsible for the content users see online (i.e., suggesting searches and order of videos). For example, Nelly Pailleux from Check First mentioned that Google represents 86% of desktop searches in Belgium. Its recommendation algorithm has no official API; thus, the platform's transparency and accountability are questionable. In Belgium, by typing "Donbas" on Google, the recommendation system suggests *The Insider* as the first position, a pro-Kremlin outlet spreading disinformation. Evidence shows that some outlets push disinformation through their recommendation system (e.g., Google, Twitter).

Much of the attention in digital policy debates is focused on content moderation, but tackling how the content is amplified is essential. For instance, TikTok blocked access to 90% of international content in Russia since the invasion of Ukraine, and Kremlin propagandists have exploited this. Shadow banning is another practice, where content reach is reduced without necessarily telling the users about it. Oppositely, there is also shadow promotion, where content that is supposed to be banned is still algorithmically promoted. Due to minimal access to algorithms, researching this tendency is very difficult. Hence, organisations spend a lot of time and resources advocating for more transparency, which is largely blocked by platforms by referring to legitimate business secrets.

Algorithm Watch audits algorithms for platforms, for instance, with the Instagram monitoring project in 2020. Adversarial audits are necessary to provide different observation scopes and to guarantee data integrity. External scrutiny is necessary because platforms cannot simply be trusted to reveal what is happening in their services. By relying on data donations and involving willing users, researchers aimed to determine whether Instagram prioritised content related to certain political groups or containing partial nudity. Findings revealed that the German right-wing party Alternative für Deutschland (AfD) and influencers' posts showing more skin appeared higher up in users' feeds. Yet, more complete data access is needed to explain these results.

When Algorithm Watch tried to get Meta (as we speak of Instagram) to comment on the results, which referred to flawed methodology, violations of GDPR and community standards, the platform announced "more formal engagement," which the Berlin-based NGO took as a threat of legal action if they continued their project. Nevertheless, the DSA grants data access to researchers vetted by regulators to study systemic risks. Moreover, the Code of Practice on Disinformation includes commitments not to take adversarial action against good-faith research and support setting up a third-party intermediary for data access requests, which needs to be specified by the European Commission in delegated acts.

Contrary to commitments in the Code of Practice, Meta is cutting down on CrowdTangle, which journalists and watchdogs relied on to monitor and report on disinformation. There is an unevenness in platforms' commitments to transparency and actions, reiterating the need for legal and procedural clarity.

PRESENTATIONS - DEBUNKING MYTHS ABOUT DISINFORMATION

How sure are we that labelling disinformation is effective? What about deplatforming? We will dive into some of our assumptions and look critically at what we know and what we are missing.

Chair: **Laura Smillie**, Policy Analyst, European Commission
Speakers: **Hella-Franziska Hoffmann**, Technical Programme Manager, Logically; **Clement Wolf**, Head of Information Quality Policy, Government Affairs and Public Policy, Google; **Prof. Stephan Lewandowsky**, Professor in Cognitive Science, University of Bristol

You can watch the recording of this session here.

Logically's mission is to reduce disinformation's harm and further its detection. The key focus is to combine human experts and artificial intelligence. The former analysed the information gathered by the latter, which detects patterns and wordings that are early indicators for influence. Thus, technology should be used to identify movements and respond with countermeasures, such as policies that imply removing content or banning accounts.

A media literacy framework encouraging cross-checking sources would make it easier for people to understand the reliability of sources. Google has more than one way of dealing with information quality and disinformation. For instance, the content advisory feature that is being rolled out warns the user about the source's reliability and relevance. Before launching such features, Google consults with media literacy experts.

The panellists reflected on the "post-truth" world in which Donald Trump was able to make 30,000 false claims throughout his presidency without the Republican party's approval ratings or number of members dropping. According to research by Hahl et al. (2018), Trump made false statements about events that could be easily disproven, thereby flagrantly flouting truth-telling norms. Nonetheless, he was perceived as an authentic champion of the "real people" against the establishment.

> *"Under a populist logic of politics, honesty and truth don't involve veracity, but they involve authenticity, belief speaking, saying things that the people want to hear. And under those circumstances lying becomes a feature and it's not a bug,"* stated Prof. Lewandowsky.

During the COVID-19 infodemic, German-speaking countries simultaneously experienced soaring trust in science and hardening vaccination hesitancy. Meanwhile, untrustworthy news outlets constitute about 14% of Facebook's engagement and 2.3% of the web overall. Although the glass is neither empty nor full, exposure to misinformation is dangerous due to its sticky nature. Corrections are often not as effective, because the "continued influence effect" kicks in.

> *"Effectiveness is still a very complicated concept to manipulate. That's because it might mean very different things to different people (...), varying according to the context,"* said Clement Wolf.

The misinformation toolbox contains no silver bullet, but there are numerous tools for debunking falsehoods which Lewandowsky et al. (2020) describe in their "Debunking Handbook." One of the most promising solutions is pre-bunking or inoculation, namely warning people that they might be misled, explaining how and why a narrative is incorrect.

In conclusion, there are explicit threats to democracy from the post-truth world. On a positive note, most people trust scientists and public health authorities. Solutions exist, so since disinformation resonates differently with different audiences, responses need to be tailored.

MASTERCLASS – THE EU DIGITAL SERVICES ACT IS A BIG DEAL FOR DISINFORMATION. HERE'S WHY

Speaker: **Rita Jonušaitė**, Advocacy Coordinator, EU DisinfoLab

You can watch the recording of this session here.

The Digital Services Act (DSA) is a ground-breaking law for internet safety and platform accountability. It differentiates between different categories of digital services: (very large) online platforms, hosting services, and intermediaries. It is mainly about illegal content, but it also demands to act on harmful content, e.g., through systemic risks mitigation and implementation of platforms' terms of service. Further key articles are Article 20 on the internal complaint-handling system, Article 21 on out-of-court dispute settlements, but also Article 86 on the representation of users' rights under the DSA by organisations.

The DSA will be helpful for disinformation experts as it will be a "data generating machine" (e.g., Articles 15, 24, 26, 39, 42, 40 for data access and scrutiny) allowing to see and investigate how platforms are complying with their obligations under the DSA. Soon, there will also be a public consultation on a dedicated delegated act on modalities for the data access for researchers, including civil society. Risk assessments (Article 34), as well as mitigation of these risks (Article 35) will require help by researchers to look into how platforms are complying. The right to lodge a complaint (Article 53) and independent audits (Article 37) will also benefit from the expertise of civil society, yet it is still unclear how auditing will exactly be done. The new Code of Practice on Disinformation will become a co-regulatory mechanism under Article 45 of the DSA. Its implementation will be critical for those platforms that will use it as their key risk mitigation measure on disinformation.

The media exemption in the DSA has been rejected in a vote by the European Parliament and not included in the final text of the legislation. This was a big win in the fight against disinformation. Still, the European Media Freedom Act proposal is paving the way for the media exemption to come back in Article 17 of the proposal! Calls for action and outreach will follow as the community needs to work together here again to make sure such dangerous ideas do not find their way back during the legislative process of the EMFA.

The DSA legislative process on the other hand has ended, and the official text entered into force 20 days after becoming available in the Official Journal of the EU. The DSA will fully apply for everyone from Q1 of 2024 onwards, but the European Commission already has information and investigative powers. In Q1 of 2023, very large online platforms and search engines will be designated, with the first risk assessments to be expected in Q2-Q3 of the same year.

In view of enforcement and accountability, the big question is whether GDPR mistakes will be repeated or learned from. Either way, the DSA is now fully adopted and here to stay.

PANEL – THE INTERNET OF TOMORROW

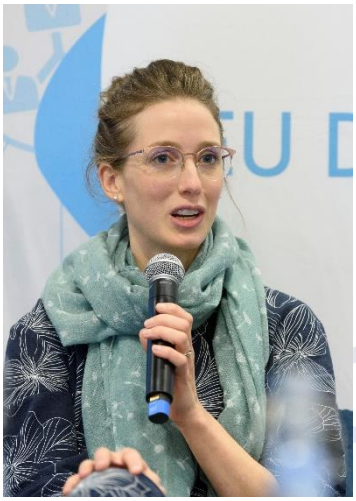
"Web3", "The Metaverse", "Synthetic Media". What do all these buzzwords actually mean, and what are their concrete implications for the fight against disinformation?

Chair: **Gregory Rohde**, Board Member, EU DisinfoLab
Speakers: **Sam Gregory**, Programme Director, Witness; **Dr. Inga Trauthig**, Senior Research Fellow, University of Texas, Austin; **Trisha Meyer**, Assistant Professor of Digital Governance and Participation, Vrije Universiteit Brussel

You can watch the recording of this session here.

The goal of the panel is to look at where we are headed and understand how the use of the internet in being transformed, what it is morphing into, and how actors use it to weaponise misinformation. The use of lies, the weaponisation of information, the attempt to exploit and divide people, to disrupt political activity is not new, but has been around for centuries What is new is we live in a time where there is an accelerant available to these bad actors that has never been seen before.

> *"Disinformation is often a manifestation of distrust in political institutions and mainstream media, and that can be easily pedalled by those who seek either financial or political profit,"* declared Prof. Trisha Meyer.



Disinformative and polarising content, such as political propaganda, widely circulates on encrypted messaging apps (e.g., WhatsApp, Telegram). Harm lies in the fact that people are more prone to trust information they receive from known contacts, and the private dimension of these conversations makes them more difficult to moderate and fact-check. Extremists such as Islamic state supporters and right-wingers exploit the deep web. In particular, the presence of Islamic state-related content is marginal on mainstream social media and hosting services, as it is pushed towards exploring alternatives than a lot of right-wing extremists, who are still successfully operating on open social media platforms. Bad actors are constantly developing online and adapting to the changing environment. A lesson to learn from this is to refrain from playing into the hype because it is directly used to undermine credible information.

"We need to start thinking that the internet of tomorrow will have more information on how media was made and that is a positive thing," said Sam Gregory.

We are moving into an internet of tomorrow where it is much easier to create mixed realities and synthetic media. It is crucial not to play into the hype and to apply a "prepare and don't panic" framework around changes in the ability to synthesise and manipulate audio, video, text, and a mix of those. There are two types of solutions for this: technical and societal. On the technical side, it should be indicated where the media comes from and how it has been edited, but still protecting privacy. The societal element should address media literacy on understanding how media is made by playing into the existing trends rather than trying to fight them. There should also be more work put in at the infrastructure level to engage with trustworthy information.

We are still looking too much at text-based solutions and not thinking about deep fakes and synthetic media in a sustainable way. Therefore, there are three main pillars to fills the gaps in disinformation responses: holding technology responsible, fact-checking, and media literacy. Disinformation will not cease to exist as a social phenomenon, and it cannot be prevented just by telling people to stop doing it. It is a malaise that needs to be listened to and not simply discredited. Moreover, we are looking at those who believe in disinformation as being passive. Instead, the core of the very problem should be addressed.