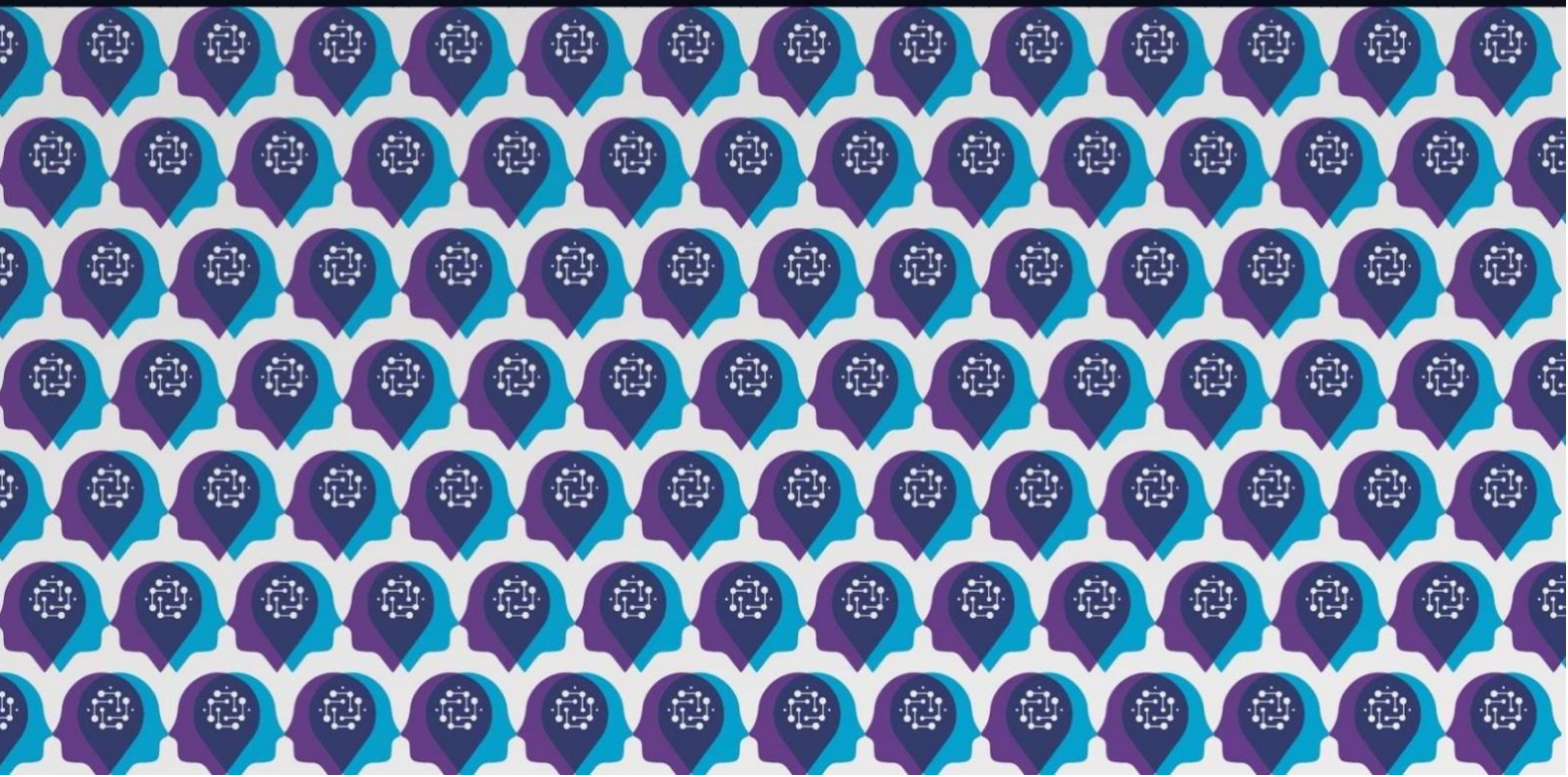
AI4Debunk

# D.12.5 : Multi-Stakeholders' perceptions

March 2025

Funded by
the European Union



Grant Agreement No.: 101135757
Call: HORIZON-CL4-2023-HUMAN-01-CNECT
Topic: HORIZON-CL4-2023-HUMAN-01-05
Type of action: HORIZON Innovation Actions

# D12.5 MULTI-STAKEHOLDERS PERSPECTIVES ON DISINFORMATION

| | |
|---|---|
| **Project Acronym** | AI4Debunk |
| **Project Number** | 101135757 |
| **Project Full Title** | Participative Assistive AI-powered Tools for Supporting Trustworthy Online Activity of Citizens and Debunking Disinformation |
| **Work package** | WP 12 |
| **Task** | Task 12.5 |
| **Due date** | 31/03/2025 |
| **Submission date** | 31/03/2025 |
| **Deliverable lead** | Partner P4D |
| **Version** | v1.0 |
| **Authors** | Gaborit Pascaline, Joen Martinsen, Vishnu Rao |
| **Contributors** | Karyna Polishchuk, Alona Hryshko, Kalina Angelova, Georgi Gotev, Zaneta Ozolina, Zane Zeibote, Marta Belsova, Sigita Struberga |
| **Reviewers** | Latvia University, CNR |
| **Abstract** | This document summarizes the findings of the stakeholders' interviews and on the 3 focus groups organized in the year 2024, in the framework of the Task 12.5. |
| **Keywords** | Stakeholders perceptions, Disinformation, War in Ukraine, Climate Change, Social Media Platforms |

Funded by the European Union

## DOCUMENT DISSEMINATION LEVEL

| Dissemination level | |
|---|---|
| X | PU – Public |
| | SEN – Sensitive |

## DOCUMENT REVISION HISTORY

| Version | Date | Description of change | List of contributor(s) |
|---|---|---|---|
| 0.1 | 01/01/2025 | | P4D, Internews Ukraine, LU, EURACTIV Bulgaria |
| 0.2 | 05/01/2025 | | LU, Internews U., EURACTIV Bulgarua |
| 0.3 | 20/01/2025 | Quality Assessment Review | CNR |
| 0.4 | 22/01/2025 | Revised version | P4D |
| 0.5 | 31/01/2025 | Project Coordinator Review | UL |
| 1.0 | 01/02/2025 | Final version ready for submission | P4D |

## STATEMENT ON MAINSTREAMING GENDER

The AI4Debunk consortium is committed to including gender and intersectionality as a transversal aspect in the project's activities. In line with EU guidelines and objectives, all partners – including the authors of this deliverable – recognise the importance of advancing gender analysis and sex-disaggregated data collection in the development of scientific research. Therefore, we commit to paying particular attention to including, monitoring, and periodically evaluating the participation of different genders in all activities developed within the project, including workshops, webinars and events but also surveys, interviews and research, in general. While applying a non-binary approach to data collection and promoting the participation of all genders in the activities, the partners will periodically reflect and inform about the limitations of their approach. Through an iterative learning process, they commit to plan and implement

Funded by the European Union

strategies that maximise the inclusion of more and more intersectional perspectives in their activities.

Funded by
the European Union

## DISCLAIMER

The AI4Debunk project has received funding from the European Union's Horizon Europe Programme under the Grant Agreement No. 101135757.

Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

## COPYRIGHT NOTICE



How to cite this report: Gaborit P. Martinsen J. Rao V. et al. 'Multi-Stakeholders Perceptions' AI4Debunk (2025). Deliverable title.

The AI4Debunk consortium is the following:

| Participant number | Participant organisation name | Short name | Country |
|---|---|---|---|
| 1 | LATVIJAS UNIVERSITATE | UL | LV |
| 2 | FREE MEDIA BULGARIA | EURACTIV | BE |
| 3 | PILOT4DEV | P4D | BE |
| 4 | INTERNEWS UKRAINE | IUA | UA |
| 5 | CONSIGLIO NAZIONALE DELLE RICERCHE | CNR-IRPPS | IT |
| 6 | UNIVERSITA DEGLI STUDI DI FIRENZE | MICC/UNIFI | IT |
| 6.1 | CONSORZIO NAZIONALE INTERUNIVERSITARIO PER LE TELECOMUNICAZIONI | CNIT | IT |
| 7 | BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION | BSC | ES |
| 8 | DOTSOFT OLOKLIROMENES EFARMOGES DIADIKTIOY KAI VASEON DEDOMENON AE | DOTSOFT | EL |
| 9 | UNIVERSITE DE MONS | UMONS | BE |
| 10 | NATIONAL UNIVERSITY OF IRELAND GALWAY | NUIG | IE |
| 11 | STICHTING HOGESCHOOL UTRECHT | HU | NL |
| 12 | STICHTING INNOVATIVE POWER | IP | NL |
| 13 | F6S NETWORK IRELAND LIMITED | F6S | IE |

Funded by the European Union

# TABLE OF CONTENTS

**1** 9**2**
112.1
12**3**
133.1
13*3.1.1*
13*3.1.2*
14*3.1.3*
14*3.1.4*
15*3.1.5*
153.2
153.3
17*3.3.1*
18*3.3.2*
19*3.3.3*
193.4
19*3.4.1*
20*3.4.2*
20*3.4.3*
21*3.4.4*
21*3.4.5*
21*3.4.6*
223.5
223.6
24**4**
27**ANNEX I. QUESTIONNAIRE FOR THE MULTI-STAKEHOLDERS INTERVIEWS AND FOCUS GROUPS** **32**

**ANNEX 2. FOCUS GROUP REPORTS** **33**

FOCUS GROUP EVENT BRUSSELS - REPORT 33
FOCUS GROUP EVENT RIGA - REPORT 35
FOCUS GROUP EVENT SOFIA - REPORT 36
FOCUS GROUP EVENT UKRAINE REPORT 38

Funded by the European Union

# ABBREVIATIONS

| | |
|---|---|
| EP | European Parliament |
| EC | European Commission |
| DSA | Digital Service Act |
| JETP | Just Energy Transition Partnership |
| PR | Public Relation |
| WP | Work Package |

Funded by the European Union

# EXECUTIVE SUMMARY

This report aims to identify key stakeholders—including media, private sector, citizens, international organizations, European public authorities, and social media platforms—and examine their roles and perceptions regarding disinformation's prevalence and impact. Through analysis of stakeholders' interviews and focus group discussions across four countries, the report explores their perspectives and gathers proposed solutions to combat disinformation. The deliverable has the following objectives:

- Identify Key Challenges: Highlight the technological, social, and institutional barriers stakeholders face in addressing disinformation.
- Capture Effective Strategies: Document existing countermeasures, including technological innovations, policy responses, and public education campaigns.
- Help partners provide recommendations: offering actionable insights for creating collaboration among stakeholders to mitigate the effects of disinformation.

By analyzing these dimensions, the report aims to contribute to ongoing global and regional efforts to combat disinformation and strengthen societal resilience against its harmful effects.
This report also supports WP5 by providing stakeholders' interviews on disinformation.

The geographic scope of this report focuses on four key locations—Brussels, Riga, Kyiv, and Sofia where the focus groups discussions have been organized.

Funded by the European Union

# 1 INTRODUCTION

Disinformation is defined by AI4DEBUNK as the deliberate creation and dissemination of false or misleading information intended to deceive. It is widely acknowledged that it has become a critical global challenge. Unlike misinformation, which is shared unknowingly, disinformation is crafted with intent, often targeting vulnerable societal structures and exploiting existing divides. Fueled by the proliferation of digital platforms and advanced technologies like artificial intelligence, disinformation campaigns now spread rapidly across borders, transcending traditional information ecosystems.

The impact of disinformation is profound, influencing democratic processes, public health, climate action, and societal trust. For instance, during recent elections worldwide disinformation campaigns were used to manipulate public opinion, polarize voters, and tarnish the reputations of political candidates. Similarly, health-related disinformation, particularly during the COVID-19 pandemic, sowed mistrust in vaccines and public health measures, while climate-related disinformation undermines global efforts to address the environmental crisis. These examples underscore the need for a comprehensive understanding of how disinformation operates and how it can be effectively countered.

The spread of misinformation and disinformation poses significant threats to public health, societal cohesion, and democratic processes.

The primary aim of this report is to identify key stakeholders' media, private sector, citizens, international organizations, public authorities in Europe, and social media platforms—regarding the prevalence and impact of disinformation, and then to analyze the multi stakeholders' interviews and focus groups discussions organized in 4 different countries to analyze the perceptions.

- Identify key challenges: Highlight the technological, social, and institutional barriers stakeholders face in addressing disinformation.
- Capture effective strategies: Document existing countermeasures, including technological innovations, policy responses, and public education campaigns.
- Help partners to provide Recommendations: Offer actionable insights for creating collaboration among stakeholders to mitigate the effects of disinformation.

By analyzing these dimensions, the report aims to contribute to ongoing global and regional efforts to combat disinformation and strengthen societal resilience against its harmful effects. The interviews included in this deliverable are instrumental in advancing the objectives of Working Package (WP) 5, particularly in relation to the identification of target groups susceptible to disinformation. By integrating insights gathered from these interviews, WP5 can leverage qualitative data to pinpoint characteristics, behaviors, and demographics of potential target audiences. This facilitates a stronger understanding of how disinformation spreads and impacts specific groups, enhancing the strategic development of counter-disinformation measures and interventions outlined within WP5 deliverables. The interviews support both this deliverable and the wider goals of WP5.

- The geographic scope of this report focuses on four key locations—Brussels, Riga, Kyiv and Sofia—where the focus group discussions have been organized. The report encompasses different stakeholder groups: media, private sector, citizens, international organizations, public authorities, and social media platforms. Not all of them were represented in the focus groups and interviews.

Funded by the European Union

- It examines disinformation in contexts such as elections, health, climate change, and the war in Ukraine.

Data collection was conducted in 2023–2024, capturing recent developments in disinformation strategies and countermeasures.

The limitations of the study described in this report are as follows:

1. Limited focus: While the findings provide valuable insights, they may not fully capture disinformation dynamics.
2. Stakeholder representation: Although diverse, the sample size only represents the view of a few stakeholders.
3. Complexity of disinformation: The rapidly evolving nature of disinformation tactics and technologies presents challenges in fully addressing their long-term impacts.

Despite these limitations, the report offers a first analysis for understanding stakeholder perceptions of disinformation and developing effective, collaborative strategies to counter its influence.

Funded by the European Union

# 2 METHODOLOGY

To obtain a comprehensive understanding of the complexities surrounding disinformation, a series of interviews were conducted with a wide spectrum of stakeholders by Pilot4dev and Internews Ukraine. These included professionals from NGOs, governmental institutions, independent journalists, activist journalists, AI tool developers, communication managers, and academics in fields such as journalism, history, security, and engineering. This diverse cohort was strategically selected to capture a holistic range of perspectives. The purpose of the research was to gain stakeholder's perceptions on the issues of disinformation and the challenges and solutions associated with it.

The interviews followed a semi-structured format, balancing adherence to a pre-defined questionnaire with flexibility to adapt questions based on the interviewee's expertise. This ensured responses were both detailed and relevant to their unique experiences. While the questionnaire still ensures some consistency and focus, making the interviews easier to compare. The questionnaire (presented in Annex 1) required the stakeholders to provide examples of evidence of disinformation with specific campaigns, describe how disinformation operates, discuss countermeasures to combat it and what measures they perceive to be most effective, and reflect on the biggest challenges and opportunities in addressing these issues, also in future prospects. The stakeholders were also asked if they had any suggestions to the AI4Debunk project itself and how they perceive AI in countering disinformation. While the interviews were guided by a fixed questionnaire, additional questions were tailored to each stakeholder's area of expertise, enabling deeper exploration of their unique insights. The questionnaire was not shared with stakeholder before the interview, to ensure that they answer the questions based on their experience and perceptions, as preparations could influence their perceptions.

The recruitment process utilized multiple strategies to identify suitable participants. An initial list of 50 potential stakeholders, derived from previous projects, served as a foundational reference, despite some outdated contact details. Further candidates were identified through independent research into organizations such as NATO StratCom, EEAS, and the International Crisis Group, as well as prominent media outlets like EuroNews and Politico. Additionally, events and webinars, including several organized by DisinfoLab and the Sofia Information Integrity Forum, were attended online to identify potential participants. Directly inviting speakers from these events significantly increased response rates, leveraging their demonstrated interest and expertise in relevant fields. Geographics were not important in the recruitment factor as suitable participants were not limited to any geographical area. Although some country diversity within Europe was desired, to collect a multitude of different perspectives within Europe on disinformation. This could especially differ significantly regarding evidence of disinformation related to the case studies. However, two stakeholders interviewed were located outside of Europe, illustrating how the recruitment was not limited to gathering European perspectives.

A total of 110 personalized e-mail invitations were sent during 2024, leading to 10 confirmed interviews. Three of these stakeholders deferred participation to 2025 due to scheduling constraints and were therefore excluded from this Deliverable. This leads to a total of seven interviews organized by Pilot4Dev, and an additional six interviews were organized by Internews Ukraine. E-mail invitations cited the

Funded by the European Union

recipient's work or recent contributions to ensure personalized outreach, enhancing the likelihood of engagement and to avoid invitations being filtered to the spam folders of the recipients.

To complement these one-on-one interviews, four focus group sessions were organized in Kyiv, Riga, Sofia, and Brussels by Internews Ukraine, The University of Latvia, Euractiv Bulgaria, and Pilot4Dev respectively. These sessions provided a platform for group discourse and debate, with each focus group gathering between 5 and 15 participants. Panel discussions addressed key themes, including participants' perceptions of disinformation, feedback on the AI4Debunk project, and innovative strategies to counter disinformation. While the focus groups followed the same questionnaire, the open structure of panel discussions led to unique debates developing in each location. The purpose of these focus groups was to facilitate experience sharing during the initial stages of the project. Being organized in the first year, the sessions focused less on presenting our work and more on gathering input and suggestions from invited stakeholders (See Annex 1 et 2).

The integration of these methodological approaches, including targeted interviews, focus groups, and participant identification through events, provided valuable insights for the AI4Debunk project, by blending expertise, practical knowledge, and collaborative insights. By aligning stakeholder engagement with the project's objectives, we gained a nuanced and multi-dimensional understanding of disinformation through this multi-stakeholder data.

## 2.1 CHALLENGES

Organizing the focus group event in Brussels presented several challenges related to participants' attendance. Initially, we had 11 confirmed participants; however, one participant informed us of their inability to attend the evening before the event due to scheduling conflicts. On the day of the event, three more participants notified us of their absence for various reasons: traffic issues while traveling to Brussels, and conflicting priorities. Despite sending reminder emails, we did not receive these notifications in time to arrange replacements for the absent participants. As a result, the number of attendees was reduced to seven, falling short of our goal of ten participants. This lower turnout somewhat affected the quality of the panel discussion, as the absence of several high-profile participants was noticeable.

Funded by the European Union

# 3 STAKEHOLDERS' PERCEPTIONS

## 3.1 EVIDENCE OF DISINFORMATION

Many stakeholders have emphasized the importance of clarifying the concepts of misinformation and disinformation. During focus group sessions, participants highlighted the need to understand the distinction between these terms. Many respondents noted that this definitional clarity was crucial when identifying evidence of disinformation (See Annex 2). At the Brussels focus group session, participants discussed how information can be misleading or manipulated without qualifying as outright disinformation. A researcher we interviewed reinforced this point, stating that "The research has focused on to what degree something could be considered disinformation. This underscores how disinformation often builds on fragments of truth, distorting them to amplify vulnerabilities and incite political polarization." Manipulated information can cause harm in various ways, and our interview findings reveal a complex landscape that includes conspiracies, deep fakes, misleading narratives, and misunderstandings.

### 3.1.1 ELECTIONS AND GOVERNMENT INSTITUTIONS

Disinformation campaigns frequently emerge during politically sensitive periods, such as elections, where actors with vested interests shape narratives to align with their political agendas. In 2024, a year marked by numerous elections, respondents observed disinformation tied closely to political motives. One respondent described "sloppy journalism" during the European Parliamentary elections, noting that media outlets often push narratives favoring specific agendas. Another respondent monitoring AI tools noted an expectation for deep fakes to surface in the lead-up to elections, underscoring the evolving nature of disinformation tactics. Although there were some deepfakes during the European Parliament elections in 2024, it is not considered to have impacted the election results according to Brad Smith from Microsoft (Mukherjee, 2024 June 3).

Governments and institutions are not immune to disinformation outside electoral periods. Such campaigns often aim to erode trust by distorting fragments of truth. One respondent noted a case in Norway, where a professor's baseless assertion that NATO was attacking Russia gained traction among audiences predisposed to distrust mainstream narratives. Professor Glenn Diesen from the University of South-Eastern Norway tweeted that "NATO will likely begin deep strikes within Russia & also destroy the Crimean bridge", adding that "NATO will claim these are Ukrainian attacks, but NATO will provide the weapons, select targets & even pull the trigger". These statements have been widely criticized as baseless claims and labeled as pro-Russian disinformation (Lohne Fosse, 2024 April 24). These efforts exploit existing skepticism of western institutions, making it challenging to counteract falsehoods and rebuilding confidence in public institutions such as NATO.

Funded by the European Union

### 3.1.2 AI AND DEEP FAKES

Technological advancements, particularly in AI and deep fakes have reshaped the disinformation landscape. Respondents pointed to generative AI models, such as deep fakes, as tools for creating and disseminating false content with ease. One respondent remarked that these tools are often used not to present plausible narratives but to evoke emotions or create “a feeling”. During the war in Ukraine, for example, both sides used satirical memes to mobilize audiences, though this content was not necessarily believed to be factual.

Respondents were divided on the current impact of AI-driven disinformation. While some noted the frequency of AI-distorted information - “at least weekly” - others argued that such content is still easily identifiable and has not yet reached the feared scale of harm. That does not mean it cannot cause harm at all. As demonstrated by the examples above, deep fakes have proven to be harmful and capable of jeopardizing women’s public persona with fake sexual content. This nuance is reflected in the respondents who expressed that deep fakes were not too harmful. “Deep fakes look really fake, but they can still be very harmful, but I don’t think it’s that big of an issue”. So, there is a recognition that deep fakes can be harmful but overall are very easy to detect.

Despite this, there was consensus about the potential risks of improving technology that could make deepfakes harder to detect in the future. “Deep fakes could erode trust in real content”, one respondent warned, adding that even genuine evidence might be dismissed as fake, undermining accountability. This sentiment, echoed by many participants, reflects growing concern about the long-term consequences of technological advancements in trust and truth.

### 3.1.3 HEALTH DISINFORMATION AND CONSPIRACY THEORIES

Health-related disinformation remains particularly damaging, with significant societal repercussions. The anti-vaccine movement, for instance, heavily relies on emotionally charged memes and distorted narratives to foster doubt about public health measures. One respondent explained how such disinformation acts as a gateway into broader conspiratorial ecosystems, gaining traction through repetition and targeted messaging. Especially on telegram, one found many channels spreading false statistics and specific anti-vaccination narratives. Another highlighted a case where disinformation campaigns funded by the Russian Embassy in Slovakia have targeted elections and COVID-19 vaccination efforts, contributing to one of the lowest vaccination rates in Europe. “The Russian Embassy in Slovakia is one of the largest in Europe... there are networks of fake news websites and profiles ,” the respondent explained. Allegations from Slovakian governments about the Russian Embassy interfering with Russian disinformation and propaganda has been denied by the embassy itself, but in 2022 a video shows a deputy military attaché from the Russian embassy bribing a contributor to the country's largest disinformation website in Slovakia (Chastand, 2022 March 24). This underscores the respondent’s argument that the embassy in Bratislava is linked to the dissemination of Russian disinformation in the country. All these

Funded by the European Union

examples illustrate how external actors systematically exploit societal vulnerabilities to deepen polarization and undermine public trust.

### 3.1.4 GEOPOLITICAL INTERESTS

Economic and geopolitical contexts also serve as fertile ground for disinformation. One respondent highlighted fake news websites in India that manipulated narratives about relationships with countries like Russia and the United States, often by copying and altering legitimate news stories such as Politico. These distortions aim to shift perceptions and sow confusion. Similarly, Russian-language media in Latvia was noted for perpetuating climate denialism, driven by Russia's broader political and fossil fuel interests. Such narratives not only distort public discourse but also undermine environmental and governmental initiatives that undermine Russian economic and geopolitical interest. Also in South Africa, there has been observed Russian influence campaigns to depict Russia as an alternative partner to the west, spreading information that they are more important in trade and military presence than they really are. These examples show how Russian influence campaigns are to be found not only in European countries, but globally with a geopolitical intention of spreading influence.

### 3.1.5 MEDIA MISREPRESENTATION AND PUBLIC PERCEPTION

Media misrepresentation, often unintentional, contributes significantly to misinformation. A respondent shared an example from South Africa where census data from Statistics South Africa was misinterpreted by media professionals, leading to widespread inaccuracies. "Fundamental misunderstandings can lead to misinformation", the respondent noted, emphasizing how even honest errors can escalate into larger disinformation narratives. It could be interesting to dive deeper into this topic with future interviews.

## 3.2 CASE STUDIES – WAR IN UKRAINE

The interviews provide extensive evidence of how disinformation about the war in Ukraine is being propagated, reflecting a sophisticated and strategic effort to shape public perception and sow division. Much of this disinformation originates from Russian state-sponsored campaigns, leveraging various platforms and tactics to achieve their goals.

In Slovakia, disinformation campaigns would be heavily influenced by proximity to Russia and the involvement of the Russian Embassy, which has been implicated in funding fake news websites and networks. These platforms disseminate narratives aimed at undermining Western alliances and spreading distrust. Similar methods are observed in Ukraine, where Russian-origin platforms like Telegram play a pivotal role in spreading anonymous and unverifiable content. Telegram, despite its popularity, becomes a critical channel for Russian propaganda targeting Russian-speaking Ukrainians and residents of southeastern Ukraine, areas directly affected by the ongoing conflict.

Funded by the European Union

Russian disinformation about the war in Ukraine often centers on narratives that question Ukraine's sovereignty and legitimacy as an independent nation. This includes framing Ukraine's government as chaotic, corrupted, and incapable of self-rule, in contrast to Russia's portrayal of itself as a stable and orderly state. These themes resonate particularly in regions with historical ties to Russia or where Russian media is consumed extensively. For example, narratives suggest that NATO and Western nations are using Ukraine as a pawn to exploit Russia's natural resources, presenting the conflict as a Western-driven conspiracy. Such claims feed skepticism about Western intentions and foster mistrust toward Ukraine's leadership, including President Zelensky.

The portrayal of the war as resource-driven is a recurring theme among the respondents we approached, with disinformation suggesting that Western powers seek to gain control over Russian gas and oil. This narrative falls into the category that the west is the aggressor in this war and that Russia is being targeted by the west. This has been amplified by online "experts" who speculate about Ukraine leveraging resource negotiations to pressure Russia or profit from the ongoing energy trade with Europe. While less prevalent within Ukraine itself, this disinformation feeds into broader regional narratives that paint the war as a contest of corrupt elites rather than a struggle for Ukrainian sovereignty. This takes away the attention from Russian aggression and war crimes in the country.

Elsewhere in Europe, similar tactics can also be reported. In France, Russian-friendly networks use subtle techniques to distort information, such as manipulating statements by French leaders like President Macron. On platforms like Telegram, this disinformation is disseminated under the guise of legitimate discourse, blurring lines between political commentary and propaganda. These campaigns exploit economic concerns, such as the costs of supporting Ukraine, to stoke public frustration and discontent with government policies. One respondent highlights that these narratives are often being pushed due to economic incentives - "When it's not done by foreign agencies, it's mostly about the money". Pointing to a trend that actors get paid to spread Russian friendly narratives and propaganda in countries like France. Globally, disinformation campaigns tied to the war also intersect AI and emerging technologies. Examples of fake websites and AI-generated newsletters have surfaced, repurposing content from legitimate outlets but embedding false narratives. This is indicative of a broader shift toward using advanced tools to craft and spread disinformation at scale, making it harder for individuals to distinguish credible information from falsehoods.

Respondents point to evidence showing that Russian disinformation uses a variety of tactics, such as exploiting local frustrations, historical conflicts, and global fears, to twist the truth about the war in Ukraine. These strategies aim not only to undermine Ukraine's fight for sovereignty but also to fracture Western solidarity and create confusion among audiences worldwide.

Marking specific campaigns, several of the interview objects had interesting cases to share. One respondent sheds light on the early role of digital platforms like Telegram in shaping the narratives surrounding the Ukraine conflict, even before the war became a global focus. It highlights how both sides utilized visual content and political messaging to mobilize support and build momentum for their respective causes. The emphasis on the uncoordinated nature of these efforts is particularly interesting,

Funded by the European Union

as it points to a decentralized or grassroots approach rather than one driven solely by state propaganda. This dynamic suggests that individuals or smaller groups, driven by personal convictions or localized interests, played an active role in crafting and disseminating these messages.

*(…) pre-war, a lot of content spreading through the telegram network, on both sides to generate support to fight in the war. A lot of visual content with political narratives. This political framing of photos was very frequently used in the beginning of the war, but also before the war broke out before most people in the rest of the world weren't paying much attention to this. These were still not coordinated campaigns. But general efforts to engage people, and this happened on both sides of the conflict.*

The research also underlines the critical prelude to the war when the conflict was built, but international audiences were largely disengaged or unaware. This gap in global attention allowed both sides to frame their narratives, setting the stage for later interpretations of the war. The early use of politically charged visuals is significant; imagery often carries emotional weight that transcends linguistic and cultural barriers, making it a powerful tool for persuasion. Even in the absence of coordinated campaigns, these efforts effectively primed both local and international audiences for the unfolding conflict.

A participant interviewed by *Internews Ukraine* also noted the use of Telegram and the fact that some of the most prevalent disinformation narratives circulate on platforms like this. They explained that these narratives commonly include claims such as power outages being caused by the government selling electricity abroad or fraudulent schemes involving "Zelensky's 1000 UAH payments". These falsehoods were propagated through Telegram, Viber chats, and Facebook posts, often they found these posts often featured AI-generated text and images. Many users, trusting the content's apparent authenticity, unwittingly shared it further.

Ultimately, these campaigns and use of social media platforms highlight the evolving nature of modern conflicts, where digital platforms like Telegram become battlegrounds for influence and engagement. It also reflects how grassroots digital campaigns can amplify political messaging, shaping perceptions long before the formal outbreak of war. This underscores the broader implications of decentralized media ecosystems in both local mobilization and global awareness.

## 3.3 CASE STUDIES – CLIMATE CHANGE

In our focus group session, we had a broader conversation about fake news and its various forms, though climate-related disinformation wasn't discussed in depth. The group recognized that false information exists on a spectrum—from unintentional misinformation to deliberate disinformation. These nuances become particularly apparent when discussing climate change. The issue is complex for several reasons: scientific evidence can appear uncertain, the concept often feels abstract to many people, and the climate crisis involves multiple sectors with competing interests. This complexity means it can be legitimate to critique certain aspects of climate change and proposed solutions. Unlike clear cases of disinformation

Funded by the European Union

(such as with the Russian invasion of Ukraine), certain patterns and claims about climate change aren't automatically disinformation. Scientific concepts can be complicated, and false climate information might simply be accidental misinformation rather than coordinated disinformation campaigns. These factors combine to make climate change disinformation particularly challenging to identify.

This complexity aligns with a recurring theme in respondent statements, where many of them were reluctant to explicitly classify climate-related skepticism or falsehoods as disinformation. Several respondents expressed uncertainty about whether these narratives stem from deliberate campaigns or simple misunderstandings. For instance, one remarked, “I wouldn’t say there’s a huge campaign behind it”, when discussing skepticism about wind turbines and climate change denialism. Others stated outright that they had not encountered any climate-related disinformation, with one commenting, “No. The closest example to disinformation of similar nature to climate change might be COVID-19 and the anti-vaccination”. Many responses were brief and vague, reflecting a perceived lack of clear evidence for organized disinformation efforts. This cautious stance underscores the complexity of defining disinformation, which often intersects with legitimate public concerns, denialism, or strategic corporate messaging. While some respondents acknowledged instances of misleading narratives, they consistently refrained from categorizing them as intentional disinformation, illustrating the broader challenge of recognizing and addressing such issues in a coherent and effective manner.

### 3.3.1 GOVERNMENT-LED INITIATIVES AND POLICIES

Despite this hesitancy, there are clear examples of climate-related disinformation, particularly targeting government-led environmental initiatives. The European Green Deal has been a frequent target of such campaigns, likely due to its ambitious and transformative goals, which have drawn significant attention and resistance. Disinformation surrounding the Green Deal has been widely documented ahead of the EP European Parliament’s elections; for instance, Dutch MEP Rob Roos claimed during a speech in Poland that the EU Green Deal was “destroying our food system” (Loucaides et al., 2024). Several of the respondents mentioned this phenomenon.

Respondents also pointed to smaller, more localized cases, such as the phase-out of plastic straws in South Africa, which sparked claims that officials were profiting from contracts for biodegradable alternatives. Similarly, conspiracy theories around climate funding mechanisms, like the Just Energy Transition Partnership (JETP), alleged that these efforts were covered for foreign exploitation. These narratives, by framing policies as politically or financially motivated, often resonate with existing skepticism toward authorities, further eroding public trust in environmental initiatives. Together, these examples demonstrate that while respondents may be hesitant to classify certain narratives as disinformation, there is clear evidence of targeted efforts to undermine trust in climate action.

Funded by the European Union

### 3.3.2 INDUSTRIAL LOBBYING

Economic interests and industrial lobbying play a significant role in shaping public perceptions of climate action. Respondents noted that industries, particularly the automotive sector, frequently employ PR (Public Relations) campaigns and spin tactics to sway opinion. One respondent highlighted encountering fake websites targeting "ethical climate activists", exemplifying how disinformation often blends with strategic corporate messaging. This deliberate manipulation blurs the line between genuine environmental concerns and industry-driven narratives, making it harder for the public to discern the truth.

One interviewee referenced an article noting that the authors of disinformation rarely cite sources, and when they do, it is never scientific literature. Instead, they often quote organizations like Clintel, a climate-skeptic group that is founded by ex-Shell employee Guus Berkhout, or the CO2 Coalition, funded for years by Koch family-linked oil foundations (Siliņš, 2023 October 3). Illustrating how climate sceptic organizations are sometimes indirectly financed by people linked to the petrol industry.

These economic pressures are sometimes intertwined with foreign influence operations. For instance, one participant noted how in Latvia, Russian-language media—shaped by Russia's broader political and fossil fuel agendas—amplifies climate denialism and skepticism, further complicating efforts to build public consensus around climate action. The participant stresses how this happens in a context where a lot of climate denialism and misunderstandings already exists in the country, making some individuals more vulnerable to Russian foreign interference on the topic.

### 3.3.3 NATURAL DISASTERS

Natural disasters and extreme weather events often become flashpoints for misinformation. While some participants acknowledged that skepticism following events like floods may not constitute organized disinformation, they noted how when Latvia saw more extreme floods than usual last year, it was followed by a lot of conspiracies as to why these floods were more extreme, that fits in a broader climate sceptic narratives. Searching for alternative explanations for why more extreme weather events happen. As one participant observed, “there are a lot of people who deny climate change exists,” pointing to the persistence of disbelief even in the face of direct environmental changes.

## 3.4 RESPONSES TO DISINFORMATION

The different interviews and focus groups enabled us to identify different responses to disinformation (See Annex 1 and Annex 2).

Funded by the European Union

### 3.4.1 PUBLIC EDUCATION

A significant emphasis is placed on public education and media literacy as essential tools, and many points to public awareness and education as the most essential measurement against disinformation. Participants agree that creating critical thinking and understanding how disinformation operates, especially regarding algorithms and AI technologies, is more impactful than reactive measures like fact-checking. Proactive education can help individuals identify and resist false narratives before they spread. One participant interviewed by Internews Ukraine stated that "It's essential for individuals to think critically to achieve results, as technology might sometimes undermine this process". This illustrates how technology is not a solution alone but requires knowledge and critical thinking from the users as well. However, media literacy is acknowledged as a long-term solution rather than an immediate fix. Short-term strategies must address current vulnerabilities, including the exploitation of low public trust and the rapid spread of disinformation via social media platforms. Many participants also acknowledge that media literacy cannot solve everything, and that although it is crucial, it needs to be implemented with other initiatives as well.

According to the respondents, educational solutions are not limited to only media literacy, but many participants perceive digital literacy just as important. Especially for the project's objective of developing AI-tools, participants perceive it as crucial that the users of these tools understand how the tool works and not trust it blindly. At the focus group in Brussels, the importance of de-mystifying machine learning, and sparking a curiosity of how AI software's works was underlined. Similarly, a participant in the Sofia focus group highlighted the importance of public understanding of AI and its mechanisms. Starting digital literacy training early in education could help foster curiosity and empower individuals to use AI tools effectively. (See Annex 2). This message was echoed in by the participants we interviewed, that digital media is crucial, also to navigate in our digital landscape today, and actually understand how the apps that surround us work. Learning about algorithms and AI-software can help people perceive these technologies seem less like magic, and through digital education can more trust be built.

Educational efforts to combat disinformation extend beyond the classroom and emphasize raising awareness among broader audiences. One key strategy, highlighted by several stakeholders, is prebunking— a proactive approach designed to expose and debunk false narratives before they gain traction. This method has been recognized as an effective way to cultivate resilience to disinformation by equipping individuals with the tools to recognize and critically assess falsehoods. The importance of prebunking was particularly underscored during the focus group session held in Kiev (Annex 2) . Participants emphasized its potential to foster critical thinking and build societal resistance to the spread of misinformation. By preemptively addressing common disinformation tactics and themes, prebunking not only counters immediate threats but also lays the foundation for a more informed and discerning public.

### 3.4.2 MEDIA FUNDING

Funded by the European Union

Media support and reform also play a vital role. Participants highlight the importance of safeguarding editorial independence and ensuring the accessibility of local and pluralistic media outlets. Trusted media is seen as a cornerstone for informing the public and countering disinformation effectively, given that people often trust media more than politicians or government agencies. Reducing the financial and bureaucratic barriers to media publication could empower new voices to enter the space, creating a more diverse information environment. Many of the participants also said that there was more funding and resources given to the media before, and that these resources have been significantly cut over the years, and that this have decreased the freedom and capabilities of the media, particularly small independent media, which is struggling in many places, and threatens the media plurality.

### 3.4.3 TRUST BUILDING

Additionally, trust was frequently brought up as an important element for building resilience against disinformation. The participants at the focus group in Kiev (Annex 2) concluded that trust building initiatives should be prioritized to counter misinformation/disinformation. This is because of how easily disinformation spreads in a distrustful environment. In one interview, a participant explained how disinformation does not appear out of nowhere, but it is able to spread where there are some uncertainties in our societies. Distrust and inequalities are making up a foundation for where disinformation can spread more easily. This was also a tip for our project on, "How are you going to act as an institution for trust" as a question we should ask ourselves when we develop a tool and to be able to actually get normal citizens. To reach out to a wider audience requires a lot of trust building, and participants perceive it as an important thing to build. However, many stressed how trust is also very hard to establish. In an interview conducted by *Internews Ukraine*, one participant noted that "(...) trust in local and regional media remains the highest. That's why it's essential to implement projects encouraging local media to create media literacy content". This recommendation highlights the importance of leveraging trusted establishments to reach people more effectively through these reliable channels.

### 3.4.4 POLICIES

Policy responses are essential, with calls for governments to institutionalize disinformation as a dedicated policy field. Measures like the EU's Digital Services Act (DSA) and the proposed international AI Act are seen as promising steps, but participants caution that legislation must be carefully crafted to avoid creating exploitable loopholes. Regulatory efforts should prioritize transparency in algorithms and require social media platforms to ensure fair visibility for reliable content. Additionally, policies should focus on enhancing funding for measures to counter disinformation, particularly in civil society, which is often outspent by well-funded disinformation campaigns supported by foreign actors.

### 3.4.5 TECHNOLOGY

Technological solutions are identified by the respondents as both a challenge and an opportunity. While AI is recognized as a tool that can enhance the spread of disinformation, it can also be harnessed to detect

Funded by the European Union

and counteract false narratives. Improved analytics, monitoring systems, and tools that enhance the reach of trustworthy information are vital. However, experts caution against over-relying on technology, emphasizing that disinformation is ultimately a societal issue that requires human oversight and judgment.

### 3.4.6 COLLABORATION

Collaboration across sectors is widely recognized as essential in the fight against disinformation. Governments, civil society organizations, media outlets, and technological platforms must join forces to address the issue in a holistic and coordinated manner. Participants emphasize that each actor brings unique strengths to the table, and leveraging these strengths is key to a successful strategy. For instance, NGOs excel in engaging with communities and creating grassroots awareness, while policymakers have valuable experience in crafting and implementing regulatory responses. To enhance these efforts, participants advocate for the establishment of centralized agencies dedicated to monitoring, analyzing, and countering disinformation. These agencies could serve as hubs for coordination, creating dialogue among various sectors, building consensus, and facilitating the exchange of best practices. Moreover, targeted awareness-raising campaigns that transparently explain the origins, methods, and goals of disinformation are vital for cultivating societal resilience. By helping citizens understand how disinformation operates, such initiatives can empower individuals and communities to recognize and resist their influence more effectively.

## 3.5 CHALLENGES AND FUTURE DIRECTIONS

The fight against disinformation faces a range of complex challenges rooted in technology, human psychology, societal inequalities, and institutional weaknesses. One of the most significant obstacles is the inability of regulations to keep pace with the rapid evolution of technology. “The issues we face now are vastly different from those of five years ago, yet our policies and regulations have not evolved to address these new realities”, noted one participant. This regulatory lag leaves society vulnerable to the unintended consequences of technological innovation, such as the proliferation of disinformation. For example, the initial enthusiasm for allowing public comments on news articles quickly gave way to concerns about unmoderated harmful discourse. One talked about comments on news articles, saying that “Moderation was introduced but often proved insufficient, and in some cases, it became better to disable comments entirely”. This illustrates how technological solutions often lag behind the challenges they create, having automatic content moderation being less sufficient than the bots generating spam content that needs to be moderated.

Disinformation, however, does not operate in a vacuum; it thrives by exploiting existing societal inequalities and insecurities. A participant emphasized, “Disinformation does not create new societal issues but exploits existing inequalities and insecurities”. In highly unequal societies like the United States, disinformation often amplifies divisions, manipulating marginalized groups or exploiting fears of a loss of status among others. This manipulation deepens distrust and fuels conflict. Addressing disinformation

Funded by the European Union

requires looking beyond the falsehoods themselves and understanding the structural vulnerabilities it amplifies. “The challenge lies not in treating disinformation as an external threat but in addressing internal vulnerabilities that it amplifies and exploits”, one contributor explained.

Human psychology also is a big challenge facing disinformation. People’s reluctance to admit they have been deceived creates a significant barrier. “Most people prefer to believe they're too smart to be tricked, and it takes a lot of courage to admit, ‘I was wrong’ or ‘I made a mistake”, one participant highlighted. This is especially true for public figures, who fear that admitting mistakes will be used against them by political rivals. Such dynamics prevent individuals and institutions from openly confronting the disinformation they have internalized or shared, leaving it to fester unchecked.

At its core, disinformation erodes trust in democratic institutions and deepens divisions in society. A participant argued that this phenomenon is not just an “information problem” but a societal one, explaining that “Disinformation becomes most corrosive when it fuels distrust in democratic institutions and divides people from their political systems—not just due to misinformation but because of deep-rooted frustrations with policy outcomes and a lack of trust”. While measures like media literacy and strategic communications are necessary, they are insufficient to address the broader discontent that makes societies susceptible to disinformation. Instead, the solution lies in strengthening democracy itself by addressing economic inequalities, providing dignity to citizens, and creating trust in institutions. “The core challenge, therefore, is to strengthen democracy itself”, a participant emphasized, underscoring the need for a holistic approach.

Independent media also plays a critical role in countering disinformation, but it faces existential threats in the form of financial pressures and suppression by platform algorithms. One participant described how small publishers are often “shadow banned” or blocked by AI tools on platforms like Meta, particularly during critical moments like election campaigns. “For example, now, during two weeks of the election campaign, all content was shadow banned and blocked because of Meta’s AI tool”, they recounted. While some publishers have resorted to newsletters to bypass these restrictions, social media remains a vital platform. “We need it to be as unfiltered as possible”, they added, emphasizing the importance of safeguarding access to these platforms to ensure the public’s right to information.

The fight against disinformation suffers from insufficient funding and limited collaboration across borders and organizations. In places like Slovakia, the effort is largely driven by small civil society initiatives that lack the resources to scale their impact. “Currently, much of the fight against disinformation is driven by small-scale civil society initiatives… which struggle to make a broader impact due to limited funding”, one expert observed. Greater investment is needed—not only in these initiatives but also in building a workforce equipped to handle the technological and strategic challenges ahead. “Investment in people is essential, especially as the field grows more complex and technologically advanced”, one participant stressed, pointing to the need for skilled professionals who can craft and implement effective policies.

There has also become a challenge that reading is less and less common, especially among younger groups. One participant interviewed by Internews Ukraine stated that it is challenging for fact-checkers to

reach out with their debunking of false news, because people are reluctant to read long-texts. Another stakeholder also noted dissemination of debunking materials as a big challenge. That text is not engaging and that their posts rarely gain the same attractions as the disinformation they are debunking.

## 3.6 RECOMMENDATIONS FOR IMPROVED COLLABORATION AMONG STAKEHOLDERS

Improved collaboration among stakeholders is vital in the fight against disinformation, and several key recommendations highlight the paths forward. A central challenge is the decline in resources for the media industry, which once played a crucial role in combating misinformation through credible, well-researched journalism. "Back then, misinformation struggled to gain traction because credible reporting dominated", one participant noted. However, declining funding has weakened the media's capacity to maintain this role. Especially for smaller and independent media outlets. Revitalizing the media requires targeted investment in fact-checking initiatives and the creation of dedicated fact-checking departments within news organizations, reminiscent of the robust research teams that once bolstered larger outlets. "Resourcing the media, particularly fact-checking initiatives, is essential", the participant emphasized, underscoring the importance of restoring the media's ability to promote accurate information. Media funding, to improve fact-checking and also pre-bunking news was brought up as important work moving forward to counter misinformation and disinformation. While, during the focus group session in Sofia (annex 2) , participants discussed the importance of collaboration between fact-checkers and journalists to create more accurate news.

There is already cooperation between fact-checkers and journalists, with most large media outlets such as the BBC and Washington Post already have their own fact-checking teams to investigate stories (Falmouth University, n.d.). Nonetheless, participants perceive it as important to maintain and expand this collaboration to better advancements in technology, particularly artificial intelligence, present opportunities to enhance fact-checking efforts, though they must be implemented carefully. One example, a fact-checking feature integrated with Alexa, inadvertently shared incorrect information due to inadequate algorithm training. "This highlights the need for substantial investment in training AI algorithms to ensure they are accurate and reliable", the same expert remarked. Leveraging AI effectively could significantly strengthen the fight against misinformation, provided it is rigorously developed and tested.

Collaboration across sectors—private, public, and governmental—is also essential. An example of successful cross-sector collaboration was cited from the digitization field, where the private sector partnered with UNICEF to connect schools to the internet. Applying a similar collaborative approach to fact-checking could yield significant progress. "We need to establish fact-checking practices and work together to identify and address common myths", one contributor noted, suggesting that pooling resources and expertise across industries could enhance the overall effort against disinformation.

Funded by the European Union

Community-level initiatives are another key aspect of collaboration. To resist disinformation effectively, campaigns must address the societal vulnerabilities that enable its spread. "Efforts should focus on understanding why people turn to alternative media and creating trust in credible sources", one participant explained. This includes creating community-driven initiatives to build local resilience against divisive issues, such as migration, LGBTQ+ rights, and climate change. By addressing these vulnerabilities at the grassroots level, stakeholders can create a more stable foundation to resist disinformation's impact.

However, political divisions often obstruct collaboration, even on critical issues. One participant observed, “Even high-priority issues, like tracking down wanted criminals, fall by the wayside when governments are at odds or lack sufficient cooperation”. They pointed to historical examples of collaboration across political and religious divides, such as during the Middle Ages and Renaissance, when cross-cultural exchanges drove progress. Today’s world needs a similar openness, but instead, political actors often exploit divisions for personal gain. "True progress will come from finding ways to connect and communicate, not from reinforcing divides”, they added, stressing the need for dialogue and unity, even between adversaries like Ukraine and Russia.

The complexity of disinformation, spanning fields from cognitive science to foreign policy, makes interdisciplinary collaboration equally essential. "Effective synergy requires realistic goals: the government should support structural synergies without overstepping its role, while civil society can offer complementary strengths", one participant explained. Fact-checking, for instance, might be better suited to media organizations, while civil society’s strengths lie in advocacy and policy development. Stakeholders should align their efforts with their core competencies to maximize impact.

Collaboration among publishers and researchers is also critical. "Researchers take the time to gather evidence, while journalists work faster and maybe more sloppily. They can complement each other", one participant suggested, emphasizing the need to bridge the gap between investigative journalism and academic research. This partnership could improve the quality of reporting and create more sustainable business models for investigative journalism.

Interdisciplinary cooperation between technical and social scientists is crucial for addressing the societal consequences of technological innovation. “Computer scientists could be harmful without talking to social scientists”, one expert warned. By creating dialogue between engineers and social scientists, the broader implications of technological inventions can be better understood and managed.

Funded by the European Union

## CONCLUSION

The findings of the AI4DEBUNK Deliverable 12.5 underscore the complexity and multidimensional nature of combating disinformation. It is evident that no single stakeholder, solution, or sector can address this challenge in isolation. Disinformation exploits gaps in technology, governance, and societal trust, thriving on existing inequalities and vulnerabilities. To counter it effectively, stakeholders must adopt a holistic, multi-pronged approach that goes beyond technical fixes to address the broader structural and societal issues that enable its spread.

Key to this effort is closing the regulatory lag that leaves societies vulnerable to technological misuse. As technology evolves rapidly, so must our policies and regulatory frameworks. Additionally, strengthening democracy through measures such as reducing inequalities, building trust in institutions, and promoting civic dignity is fundamental. Disinformation is not just an information problem, it is a societal one, deeply rooted in the erosion of trust and widening socio-political divisions.

Collaboration emerges as a recurring theme and a critical requirement for success. Whether it is cross-sector partnerships between governments, private entities, and civil society, or interdisciplinary cooperation between technical and social sciences, collective efforts are essential. Initiatives must draw on the complementary strengths of different stakeholders, leveraging expertise, resources, and reaching to develop innovative and sustainable solutions.

The role of the media, both traditional and new, cannot be overstated. Revitalizing independent journalism and investing in fact-checking initiatives are crucial to maintaining the integrity of information. Equally important is addressing the systemic challenges posed by platform algorithms, ensuring that they support rather than hinder access to accurate information. Emerging technologies such as AI offer potential solutions but require careful development and rigorous oversight to avoid exacerbating the problem they aim to solve.

Ultimately, the fight against disinformation must prioritize building societal resilience. This involves fostering media literacy, addressing underlying vulnerabilities, and engaging communities in dialogue to counter divisive narratives. By empowering citizens and fostering trust in credible information sources, stakeholders can create a more informed, connected, and resilient society.

The path forward requires sustained effort, significant investment, and an unwavering commitment to cooperation. As disinformation continues to evolve, so too must our strategies, guided by a shared vision of truth, equity, and democratic integrity. Only through collective action can we mitigate the corrosive effects of disinformation and build a foundation for a more informed and united world.

Funded by the European Union

# 4 REFERENCES


Andersson, L. (2021). It's critical: The role of critical thinking in media and information literacy. Media Education Research Journal, 10(1&2), 1–14. https://microsites.bournemouth.ac.uk/merj/files/2021/12/MERJ-10.12-Andersson.pdf

ARTICLE 19. (2022, April 5). UN: Human Rights Council adopts resolution on disinformation. Retrieved December 18, 2024, from https://www.article19.org/resources/un-human-rights-council-adopts-resolution-on-disinformation/

Arcuri, M. C., Gandolfi, G., & Russo, I. (2023). Does fake news impact stock returns? Evidence from US and EU stock markets. Journal of Economics and Business, 125, 106130. https://www.sciencedirect.com/science/article/pii/S0148619523000231

Association for Progressive Communications. (2021, February). Disinformation and freedom of expression. Retrieved December 20, 2024, from https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/APC-Disinformation-Submission.pdf

Ask EP. (2022, April 21). EU action against 'fake news'. European Parliamentary Research Service Blog. Retrieved December 19, 2024, from https://epthinktank.eu/2022/04/21/eu-action-against-fake-news/

Bank of America. (n.d.). The threat misinformation and disinformation pose to business. Cyber Security Journal, Issue 7. Retrieved December 18, 2024, from https://business.bofa.com/content/dam/flagship/global-transaction-services/misinformation/Misinformation.pdf

Bantourakis, M. (2023, October). News media, literacy, and trust in the age of AI. World Economic Forum. https://www.weforum.org/stories/2023/10/news-media-literacy-trust-ai/

Barrett, N. (2024, October 12). Facebook, X and TikTok: How social media algorithms shape speech. BBC News. Retrieved December 19, 2024, from https://www.bbc.com/news/articles/cp8e4p4z97eo

Basuroy, T. (2024, March 22). India: Frequency of being misled by fake news online 2022. Statista. Retrieved December 20, 2024, from https://www.statista.com/statistics/1388664/india-frequency-of-being-misled-by-fake-news-online/

Bateman, J., & Jackson, D. (2024). Countering disinformation effectively: An evidence-based policy guide. Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en

BBC. (n.d.). Faktisk.no: Norway's pioneering fact-checking organisation. Beyond Fake News: Trusted News Initiative. Retrieved December 18, 2024, from https://www.bbc.com/beyondfakenews/trusted-news-initiative/faktisk-no-norways-pioneering-fact-checking-organisation

Berger, C., & Unzicker, K. (2024, February 28). Vast majority views disinformation as a threat to democracy and social cohesion. Bertelsmann Stiftung. Retrieved December 20, 2024, from https://www.bertelsmann-stiftung.de/en/topics/latest-news/2024/february/vast-majority-views-disinformation-as-a-threat-to-democracy-and-social-cohesion

Bulger, M., & Davison, P. (2018). The promises, challenges, and futures of media literacy. Journal of Media Literacy Education, 10(1), 1-21. https://digitalcommons.uri.edu/cgi/viewcontent.cgi?params=%2Fcontext%2Fjmle%2Farticle%2F1365%2F&path_info=Bulger_and_Davison_JMLE_10.1.pdf

· Cabrera Blázquez, F. J. (2021). European Democracy Action Plan. IRIS Legal Observations of the European Audiovisual Observatory, 2021(2), Article 1/4. Retrieved December 19, 2024, from https://merlin.obs.coe.int/article/9099

Castro, M. C. (2024, October 15). Fake news, corruption and compliance in the private sector. Transparency International. Retrieved December 18, 2024, from https://knowledgehub.transparency.org/assets/uploads/helpdesk/Fake-news-corruption-and-compliance-in-the-private-sector.pdf

Cathy. (2023, October 4). How misinformation and disinformation impact global business? Bocasay. Retrieved December 18, 2024, from https://www.bocasay.com/misinformation-disinformation-impact-global-business/

Chandran, R., & Galuh, L. (2022, November 21). In Asia, citizen groups fill fact-checking gap left by Big Tech. The Christian Science Monitor. Retrieved December 20, 2024, from https://www.csmonitor.com/Technology/2022/1121/In-Asia-citizen-groups-fill-fact-checking-gap-left-by-Big-Tech

Chastand, J.-B. (2022, March 24). In Slovakia, the disinformant was paid by Moscow. *Le Monde*. Retrieved from https://www.lemonde.fr/en/international/article/2022/03/24/in-slovakia-the-disinformant-was-paid-by-moscow_5979091_4.html

Clune, C. (2023, June 19). Behind the content moderation strategies of social media giants. UNSW BusinessThink. Retrieved December 19, 2024, from https://www.businessthink.unsw.edu.au/articles/content-moderation-strategies-social-media

Damiani, J. (2019, September 3). A voice deepfake was used to scam a CEO out of $243,000. Forbes. Retrieved December 18, 2024, from https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

Dame Adjin-Tettey, T. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. Cogent arts & humanities, 9(1), 2037229. https://www.tandfonline.com/doi/pdf/10.1080/23311983.2022.2037229

· Davis, S. (2024, February 29). New study unveils strategies to combat disinformation wars on social media. Tepper School of Business, Carnegie Mellon University. Retrieved December 19, 2024, from https://www.cmu.edu/tepper/news/stories/2024/march/disinformation-wars-combat-strategies-social-media.html

De Gregorio, G., & Pollicino, O. (2024, June). The European approach to disinformation: Policy perspectives (Policy Brief No. 19). Institute for European Policymaking @ Bocconi University. Retrieved December 19, 2024, from https://iep.unibocconi.eu/sites/default/files/media/attach/PB19_Disinformation_%20Pollicino.pdf

Di Domenico, G., & Ding, Y. (2023). Between brand attacks and broader narratives: How direct and indirect misinformation erode consumer trust. Current Opinion in Psychology, 54, 101716. https://doi.org/10.1016/j.copsyc.2023.101716

Dame Adjin-Tettey, T. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. Cogent arts & humanities, 9(1), 2037229. https://www.tandfonline.com/doi/pdf/10.1080/23311983.2022.2037229

E-SPIN Group. (2024, December 6). Disinformation security in business: Mitigating the negative impacts. Retrieved December 18, 2024, from https://www.e-spincorp.com/disinformation-security-in-business-mitigating-the-negative-impacts/

Espaliú Berdud, C. (2024). The EU Code of Practice on Disinformation: An example of the self-regulatory trend in international and European law. VISUAL Review, 16(2), 95–109. Retrieved December 19, 2024, from https://visualcompublications.es/revVISUAL/article/download/5217/3519/17919

European Commission. (n.d.). Media literacy. Retrieved December 19, 2024, from https://digital-strategy.ec.europa.eu/en/policies/media-literacy

European Commission. (n.d.). Online disinformation. Retrieved December 19, 2024, from https://digital-strategy.ec.europa.eu/en/policies/online-disinformation

European Court of Auditors. (2020, March). EU action plan against disinformation: Audit preview. Retrieved December 20, 2024, from https://www.eca.europa.eu/lists/ecadocuments/ap20_04/ap_disinformation_en.pdf

European Education and Culture Executive Agency. (n.d.). Media literacy and safe use of new media. Youth Wiki. Retrieved December 19, 2024, from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/bulgaria/68-media-literacy-and-safe-use-of-new-media

European External Action Service. (2019, March). Rapid Alert System: Strengthening coordinated and joint responses to disinformation [Factsheet]. Retrieved December 19, 2024, from https://www.eeas.europa.eu/sites/default/files/ras_factsheet_march_2019_0.pdf

European Union. (2024). Mapping of the media literacy initiatives, landscape, and industry in the context of Better Internet for Kids. European Union. Retrieved December 19, 2024, from https://media-and-learning.eu/type/news/overview-of-media-literacy-initiatives-in-eu-available/

Falmouth University. (n.d.). 5 ways journalists fact-check stories in the age of fake news. Retrieved December 12, 2024, from https://www.falmouth.ac.uk/news/5-ways-journalists-fact-check-stories-fake-news

Flore, M., Balahur, A., Podavini, A., & Verile, M. (2019). Understanding citizens' vulnerabilities to disinformation and data-driven propaganda. Publications Office of the European Union. Retrieved December 20, 2024, from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC116009/understanding_citizens_vulnerabilities_to_disinformation.pdf

Greškovičová, K., Masaryk, R., Synak, N., & Čavojová, V. (2022). Superlatives, clickbaits, appeals to authority, poor grammar, or boldface: Is editorial style related to the credibility of online health messages? Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.940903

Haro-de-Rosario, A., Sáez-Martín, A., & Caba-Pérez, M. del C. (2018). Using social media to enhance citizen engagement with local government: Twitter or Facebook? New Media & Society, 20(1), 29–49. doi: 10.1177/1461444816645652

Hetler, A. (2024, February 8). 6 content moderation guidelines to consider. TechTarget. Retrieved December 19, 2024, from https://www.techtarget.com/whatis/feature/Content-moderation-guidelines-to-consider

Human Rights Watch (2024, May 6). South Africa: Toxic Rhetoric Endangers Migrants. *Humans Rights Watch*. Retrived from https://www.hrw.org/news/2024/05/06/south-africa-toxic-rhetoric-endangers-migrants

Hunt, S., & Overby, S. (n.d.). Misinformation in business: Don't eradicate, educate. SAP Insights. Retrieved December 18, 2024, from https://www.sap.com/india/insights/viewpoints/fighting-misinformation-in-business.html

Ignatidou, S. (2019, October 18). EU-US cooperation in tackling disinformation: Countermeasures to disinformation. Chatham House. Retrieved December 20, 2024, from

https://www.chathamhouse.org/2019/10/eu-us-cooperation-tackling-disinformation-0/3-countermeasures-disinformation

International Fact-Checking Network. (n.d.). International Fact-Checking Network. Poynter Institute. Retrieved December 19, 2024, from https://www.poynter.org/ifcn/

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. American Behavioral Scientist, 65(2), 371–388. https://doi.org/10.1177/0002764219869406

Kellner, D., & Share, J. (2007). Critical media literacy: Crucial policy choices for a twenty-first-century democracy. Policy Futures in Education, 5(1), 59-69. https://journals.sagepub.com/doi/pdf/10.2304/pfie.2007.5.1.59

Kőváry Sólymos, K. (2023, October 31). Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse. *International Press Institute (IPI)*. Retrieved from https://ipi.media/slovakia-deepfake-audio-of-dennik-n-journalist-offers-worrying-example-of-ai-abuse/

Kumar, B., & Rosenbach, E. (2020, July). Deepfakes and synthetic media in the financial system: Assessing threat scenarios. Carnegie Endowment for International Peace. Retrieved December 18, 2024, from https://carnegieendowment.org/research/2020/07/deepfakes-and-synthetic-media-in-the-financial-system-assessing-threat-scenarios?lang=en

Lohne Fosse, A. (2024, April 24). Norsk professor vekker reaksjoner: NATO vil angripe i Russland. *Nettavisen*. Retrieved from https://www.nettavisen.no/nyheter/norsk-professor-vekker-reaksjoner-nato-vil-angripe-i-russland/s/5-95-1780596

Loucaides, D., Edwards, S., & Perrone, A. (2024, April 30). Conspiracy theorists have turned from COVID to climate: How will it impact the EU elections? *Euronews*. Retrieved from https://www.euronews.com/green/2024/04/30/conspiracy-theorists-have-turned-from-covid-to-climate-how-will-it-impact-the-eu-elections

López-Marcos, C., & Vicente-Fernández, P. (2021). Fact Checkers Facing Fake News and Disinformation in the Digital Age: A Comparative Analysis between Spain and United Kingdom. Publications, 9(3), 36. https://doi.org/10.3390/publications9030036

Ma, O., & Feldman, B. (2022, November 29). How Google and YouTube are investing in fact-checking. Google Blog. Retrieved December 18, 2024, from https://blog.google/outreach-initiatives/google-news-initiative/how-google-and-youtube-are-investing-in-fact-checking/

Markowitz, D. M., Levine, T. R., Serota, K. B., & Moore, A. D. (2023). Cross-checking journalistic fact-checkers: The role of sampling and scaling in interpreting false and misleading statements. PLOS One, 18(7), e0289004. https://doi.org/10.1371/journal.pone.0289004

Martens, H., & Hobbs, R. (2015). How media literacy supports civic engagement in a digital age. Atlantic Journal of Communication, 23(2), 120-137. https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1048&context=com_facpubs

Mishra, A. (2024, April 1). Meta partners with PTI to expand its fact-checking program in India. Business Standard. Retrieved December 19, 2024, from https://www.business-standard.com/india-news/meta-partners-with-pti-to-expand-its-fact-checking-program-in-india-124040100728_1.html

National Center for State Courts. (n.d.). Disinformation and the public. Retrieved December 20, 2024, from https://www.ncsc.org/consulting-and-research/areas-of-expertise/communications%2C-civics-and-disinformation/disinformation/for-the-public

Nielsen, R., & Graves, L. (2017). "News you don't believe": Audience perspectives on fake news. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/public-perspectives-trust-news

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020, June 2). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. HKS Misinformation Review. Retrieved December 20, 2024, from https://misinforeview.hks.harvard.edu/article/misinformation-in-action-fake-news-exposure-is-linked-to-lower-trust-in-media-higher-trust-in-government-when-your-side-is-in-power/

Park, S., Lee, J. Y., Notley, T., & Dezuanni, M. (2023). Exploring the relationship between media literacy, online interaction, and civic engagement. The Information Society, 39(4), 250–261. https://doi.org/10.1080/01972243.2023.2211055

Pew Charitable Trusts. (2024, October). Media mistrust has been growing for decades—Does it matter? Retrieved December 18, 2024, from https://www.pewtrusts.org/en/trend/archive/fall-2024/media-mistrust-has-been-growing-for-decades-does-it-matter

Prevency. (n.d.). Don't let disinformation destroy your brand: Tips for companies to identify and combat the threat of ChatGPT and other AI tools. Retrieved December 18, 2024, from https://prevency.com/en/dont-let-disinformation-destroy-your-brand-tips-for-companies-to-identify-and-combat-the-threat-of-chatgpt-and-other-ai-tools/

PricewaterhouseCoopers. (2021, February 9). Disinformation attacks in the corporate sector: Are you ready? PwC. https://www.pwc.com/us/en/tech-effect/cybersecurity/corporate-sector-disinformation.html

PressReader. (n.d.). Media literacy and civic engagement: Empowering citizens to participate in democratic discourse. Retrieved December 19, 2024, from https://blog.pressreader.com/libraries-institutions/media-literacy-and-civic-engagement-empowering-citizens-to-participate-in-democratic-discourse

Putnam, R. D. (1996). The strange disappearance of civic America. Policy: A Journal of Public Policy and Ideas, 12(1), 3-15. http://ereserve.library.utah.edu/Annual/POLS/3010/Button/strange.pdf

Raghunath, D., & Malik, S. (2024, March 1). Partnering with news publishers and fact-checkers to tackle misinformation ahead of the Indian General Elections 2024. The Keyword. https://blog.google/intl/en-in/company-news/partnering-with-news-publishers-and-fact-checkers-to-tackle-misinformation-ahead-of-the-indian-general-elections-2024/

Romero, R. (2024, October 27). How social media algorithms are evolving to combat fake news. Multipost Digital. Retrieved December 19, 2024, from https://www.multipostdigital.com/blog/xpe0bfj7bjgwwpcxmf9vnsd35n37vr

Salgar, S. (2024, March 15). A comprehensive approach to content moderation in social media platforms. INDIAai. Retrieved December 19, 2024, from https://indiaai.gov.in/article/a-comprehensive-approach-to-content-moderation-in-social-media-platforms

Sell, T. K., Hosangadi, D., Smith, E., Trotochaud, M., Vasudevan, P., Gronvall, G. K., Rivera, Y., Sutton, J., Ruiz, A., & Cicero, A. (2021, March). National priorities to combat misinformation and disinformation for COVID-19 and future public health threats: A call for a national strategy. Johns Hopkins Center for Health Security. Retrieved December 20, 2024, from https://centerforhealthsecurity.org/sites/default/files/2023-02/210322-misinformation.pdf

Funded by the European Union

Serrano, J. (2024, March 7). Disinformation is a threat to our trust ecosystem. Experts explain how to curb it. World Economic Forum. Retrieved December 20, 2024, from https://www.weforum.org/stories/2024/03/disinformation-trust-ecosystem-experts-curb-it/

Siliņš, R. (2023 October 3). Pants on fire – how climate misinformation is spread in the Baltic states?. *Re:BALTICA*. Retrived from https://en.rebaltica.lv/2023/10/pants-on-fire-how-climate-misinformation-is-spread-in-the-baltic-states/

Spratt, H. E., & Agosto, D. E. (2017). Fighting fake news: Because we all deserve the truth. Young Adult Library Services, 15(4), 17–21. https://webreeldata.s3.amazonaws.com/449_Fighting%20Fake%20News-%20Becasue%20We%20All%20Deserve%20the%20Truth.pdf

Subramanian, S. (2017, February 15). Inside the Macedonian fake-news complex. WIRED. https://www.wired.com/2017/02/veles-macedonia-fake-news/

Supantha Mukherjee (2024, June 3). Few AI deepfakes identified in EU elections, Microsoft president says. Reuters, retrived from https://www.reuters.com/technology/few-ai-deepfakes-identified-eu-elections-microsoft-president-says-2024-06-03/

Thabi Myeni. (2022, June 16). S African minister suggests foreigners are to blame for no jobs. Aljazeera, Retrived from https://www.aljazeera.com/news/2022/6/16/s-africa-minister-suggests-foreigners-to-blame-for-no-jobs

United Nations. (2022, August 12). Countering disinformation for the promotion and protection of human rights and fundamental freedoms: Report of the Secretary-General (A/77/287). Retrieved December 18, 2024, from https://documents.un.org/doc/undoc/gen/n22/459/24/pdf/n2245924.pdf

UNESCO. (2022, March 30). Freedom of expression, disinformation and democracy debated at UNESCO Brainboost session. Retrieved December 18, 2024, from https://www.unesco.org/en/articles/freedom-expression-disinformation-and-democracy-debated-unesco-brainboost-session

Upton, P., Kashifuddin, M., Joyce, S., & Nocera, J. (2021, February 9). Disinformation attacks in the corporate sector: Are you ready? PwC. https://www.pwc.com/us/en/tech-effect/cybersecurity/corporate-sector-disinformation.html

Ward, B. (2022, April 13). Beware misinformation and propaganda about the cost of energy crisis. Grantham Research Institute on Climate Change and the Environment. Retrieved December 18, 2024, from https://www.lse.ac.uk/granthaminstitute/news/beware-misinformation-and-propaganda-about-the-cost-of-energy-crisis/

Winter, K., Hornsey, M. J., Pummerer, L., & others. (2024). Public agreement with misinformation about wind farms. Nature Communications, 15, 8888. https://doi.org/10.1038/s41467-024-53278-2

Young European Ambassadors. (2023, May 31). Cooperation between the EU and Ukraine to fight disinformation during wartime. EU NEIGHBOURS east. Retrieved December 19, 2024, from https://euneighbourseast.eu/young-european-ambassadors/blog/blog-cooperation-between-the-eu-and-ukraine-to-fight-disinformation-during-wartime/

Funded by the European Union

Funded by
the European Union

# ANNEX 1. QUESTIONNAIRE FOR THE MULTI-STAKEHOLDERS INTERVIEWS AND FOCUS GROUPS

**Questionnaire Stakeholder AI4 DEBUNK**

Stakeholder Questionnaire: Addressing Disinformation

This questionnaire is designed to gather insights from stakeholders involved in the fight against disinformation. This questionnaire aims to understand the current challenges, effective responses, and evidence of interference in the spread of disinformation and fake news:

**Personal Information:**

**Full Name:**

- Please provide your full name

**Position:**

- What is your current role and in which organization?

**Evidence of Disinformation:**

- Have you observed any direct evidence of disinformation within your area of responsibility?: If yes, please describe the evidence and the context in which it was observed
- Can you identify any specific sources or campaigns of disinformation that have had a significant impact? Please describe the sources, nature of the campaigns, and their impacts

**Case Studies:**

*Disinformation Related to Climate Change*:

- Have you encountered specific instances of disinformation regarding climate change? (Please provide details about the nature, source, and impact of the disinformation.)
- What measures have been effective in countering this type of disinformation? (Describe any strategies or initiatives undertaken specifically to address climate change disinformation.)

*Disinformation Related to the War in Ukraine:*

- Have you encountered specific instances of disinformation regarding the war in Ukraine? (Please provide details about the nature, source, and impact of the disinformation.)
- What measures have been effective in countering this type of disinformation? (Describe any strategies or initiatives undertaken specifically to address disinformation related to the war in Ukraine.)

**Responses to Disinformation:**

- What measures has your organization implemented to combat disinformation and fake news? (Please list and describe the measures taken.)
- How do you assess the effectiveness of these measures? (Describe any metrics or methods used to evaluate the effectiveness.)
- In your view, what are the most effective types of responses against disinformation?

Policy changes

Public education and awareness

Technological solutions

Collaboration with other entities

Funded by the European Union

Other (please specify)
(Please elaborate on why you believe these responses are the most effective.)

**Challenges and Future Directions:**

- What are the biggest challenges you face in combating disinformation? (Describe the challenges, including any resource, policy, or technological barriers.)
- What additional resources or changes would help your organization more effectively fight disinformation? (Specify any resources, support from government or other organizations, policy changes, etc.)
- Do you have any recommendations for improving collaboration among stakeholders in the fight against disinformation? (Suggestions for enhanced coordination, sharing best practices, joint initiatives, etc.)

**Additional Comments:**

- Please provide any additional comments or insights you wish to share about the fight against disinformation.

# • ANNEX 2. FOCUS GROUP REPORTS

This annex contains the short reports from the 4 focus group events organized in Brussels, Riga, Sofia and Kiev.

## ○ FOCUS GROUP EVENT BRUSSELS - REPORT

**Project:** AI4Debunk
**Date**: 5th of September
**Meeting Outline**

· Welcoming and networking talks
· Introductory presentation
· Round-table presentations
· Project Overview and Expectations
· Questions and discussion

The meeting aimed to exchange information, avoid redundancies, and understand the efforts of various organizations. The AI4Debunk project and the first findings of the online poll were presented. It was followed by a Round-Table introduction, and a discussion on the potential risks, including duplicating existing work, and the need for high-quality information. 7 participants gathered including the team.

**Discussion**

Following the presentation, the participants initiated a lively discussion and raised numerous questions. This conversation used a semi flexible list of questions . The discussion highlighted not only the societal

Funded by the European Union

aspects of disinformation but also the individual psychological factors that contribute to someone's vulnerability to being exposed to and believing manipulated information online. It was suggested that these psychological elements could be considered when designing the tools aimed at combating disinformation.

Participants expressed interest in the design of the tools, questioning why these four specific interfaces (Web plug in, AR/VR, App, Collaborative Platform ) were chosen and how they aligned with the project's goals. There were equal concerns about making the tools accessible to the general public, especially those who might benefit the most from them, such as seniors or less tech-savvy individuals. Although there was some skepticism regarding whether these groups would readily adopt apps or interfaces like VR/AR systems, the general consensus was that the tools, especially the app, could be designed to be user-friendly.

A major focus of the discussion was on how to encourage widespread use of these tools, specifically in building trust as a credible source. Several stakeholders acknowledged the challenge of establishing trust, as even recognized fact-checkers face difficulties in gaining the public's confidence. Vulnerable individuals, in particular, were seen as less likely to trust AI-driven fact-checking tools, while more informed users might be willing to experiment with them.

The participants also inquired about our definitions of "disinformation" and "misinformation." This led to a broader discussion on the concept of fake news and its various shades, acknowledging that false information exists on a spectrum ranging from unintentional misinformation to deliberate disinformation. The way information is framed, especially through the use of emotional language, was seen as influencing how readers interpret content. Some mentioned that AI systems already exist that can assess the emotional tone of articles, yet concerns were raised about the risk of overgeneralizing different types of false information. The complexity of determining truth is why fact-checkers continue to rely on human analysis—an approach also central to our Disinfopedia. There was debate over whether established media outlets should be notified when they publish incorrect information, with some advocating for a collaborative effort between journalists and fact-checkers to ensure the accuracy of news. Transparency from journalists when they make mistakes was viewed as a possibility to maintaining public trust.

The discussion shifted towards the issue of malicious actors and their deliberate efforts to spread false information. Identifying these actors and tracing the origins of such threats is often incredibly difficult. The DisinfoLab team, which has explored this challenge extensively, shared their published findings available in this paper:
https://www.veraai.eu/posts/report-revisit-coordinated-inauthentic-behaviour-detection-tree

Participants were curious about our own experiences in tracking malicious actors. Drawing from our case study, we highlighted how Russian disinformation tactics have proven to be particularly adaptive. For instance, their strategy on platforms like TikTok focuses on attracting younger, liberal audiences with engaging content, gradually introducing propaganda after first building trust and engagement. This approach of embedding misinformation in narratives that resonate with specific demographics poses a

Funded by the European Union

growing threat. It is especially concerning on platforms with weaker disinformation controls, like TikTok, compared to platforms such as Meta or X (formerly Twitter).

Another important point raised was the need to not rely blindly on AI when countering fake news but to encourage curiosity and understanding of how AI systems work. Some participants called for efforts to "de-mystify AI" and large language models, arguing that people often passively engage with AI without understanding its underlying processes. They suggested that fostering curiosity about machine learning from an early age could help in making these tools more accessible to the general public. This was part of the rationale behind the introduction of a comic book as an educational tool in the project, sparking a discussion about whether educating people on AI's role in fighting disinformation could lead to wider use of these tools beyond just professionals.

Towards the end of the discussion, participants referred to their experiences with the VeraAI project, which has been running for three years. They saw potential for future collaboration between this ongoing project and the current initiative, suggesting that there could be valuable connections to explore moving forward.

## ○ FOCUS GROUP EVENT RIGA - REPORT

On November 1, 2024, a multi-stakeholder meeting was held at the House of Europe in Riga, Latvia from 12:00 to 13:30. The session gathered 12 representatives from various sectors, including government ministries, the business community, research institutions, and non-governmental organizations.
Professor Inna Šteinbuka, the project lead, opened the discussion by presenting the AI4Debunk project. She highlighted the importance of regular coordination and cooperation among all stakeholders engaged in addressing artificial intelligence (AI) and disinformation. She emphasized that broad dissemination of the project's outcomes would be critical, with stakeholder networks playing an essential role in ensuring these results reach the necessary audiences.

A central theme of the discussion was the recognition of disinformation as a "new normal." Participants observed that, particularly in the Baltic States, the prevalence, influence, and impact of disinformation are well understood. Concerns were raised regarding increasing foreign interference in European politics, notably from Russia and China, which poses significant challenges to regional stability.

Regarding AI's role, attendees generally agreed that AI alone cannot fully address the complexities of disinformation. A more comprehensive and systemic approach is needed, with AI tools tailored to the requirements of specific target groups. For instance, media outlets and social media platforms might deploy AI to support fact-checking and disinformation tracking. Participants emphasized the importance of identifying the distinct needs of these groups before designing AI tools.

Funded by the European Union

The conversation also touched on regulatory mechanisms. While participants commended the adoption of the EU AI Act, they raised questions about its implementation and practical outcomes. There was broad support for creating a transparent directory of media entities, detailing ownership, beneficiaries, and funding sources, which would assist in tracing disinformation.

Trust in technology emerged as a key issue. Survey data indicates a declining public trust in advanced technological tools. This skepticism toward AI's capability to detect and curb disinformation calls into question the effectiveness and acceptance of tools to be elaborated at the end of the AI4Deunk project.
Challenges in implementing AI-based tools to combat disinformation were also discussed. A significant concern was the ongoing maintenance, upgrading, and training required for these tools in the post-project period. Participants emphasized the importance of the "human-in-the-loop" approach, which addresses the need for ethical decision-making, media literacy, an adaptive education system, and lifelong learning initiatives.

The role of NGOs in countering disinformation was examined from multiple perspectives. Latvia has a robust network of NGOs focusing on ICT issues, which could be instrumental in disseminating the AI4Debunk project results. In collaboration with the project team, NGOs could develop tailored information packages for different audiences. Moreover, NGOs could offer valuable insights into the specific needs of various target groups regarding AI solutions.
Overall, this multi-stakeholder meeting proved to be highly constructive. It served as a foundation for ongoing collaboration and potential future partnerships to effectively disseminate the outcomes of the AI4Debunk project.

### ○ FOCUS GROUP EVENT SOFIA - REPORT

*Club 'Journalist' at the Union of Bulgarian Journalists, Sofia*

**Agenda**

1. **Welcoming and Networking**
2. **Introductory Presentation**
3. **Project Overview and Expectations**
4. **Audience Q&A and Discussion**

The meeting served as a platform to share insights and explore the challenges of combating fake news and disinformation through artificial intelligence. The presentation offered the audience a comprehensive overview of the AI4Debunk project, covering its structure, objectives, and the consortium behind it. It detailed the progress on the tool's development, shed light on the contributions of Euractiv Bulgaria, and on key insights and milestones achieved so far. An open discussion on disinformation and the role of the media, and expectations for the project followed. It was attended by 11 participants, including the project team, and aimed to exchange ideas, highlight diverse perspectives, and refine the development of tools against disinformation.

Funded by the European Union

- **Discussion Highlights**
- **Audience Questions and Responses**

Following the presentation, participants engaged in an active discussion using a semi-structured format. Audience feedback provided valuable insights into the perception of disinformation and the potential applications of AI tools in addressing it:

- **Familiarity with the Project**: Five participants were already aware of *AI4Debunk*, while three encountered it for the first time.
- **Perceptions of Fake News**: Six attendees viewed fake news as a serious threat, but one considered the topic exaggerated.
- **Experience with Artificial Intelligence**: Seven participants had used AI tools, whereas three had not.
- **Encounters with AI-Generated Fake News**: Six participants had come across AI-generated fake news, though one admitted they paid little attention to it.
- **Combatting Disinformation**: Seven agreed on the importance of addressing fake news, while one felt the issue was overblown.
- **Bulgaria's Vulnerability**: Nine participants believed that Bulgaria is particularly susceptible to disinformation campaigns.
- **Willingness to Participate**: Seven expressed interest in joining projects like *AI4Debunk*, while one showed no interest.
- **Personal Experience with Fake News**: Five admitted falling for fake news in the past, and eight observed fake news in established media outlets.
- **Efforts to Protect Others**: Nine participants reported warning family and friends about disinformation.
- **Ability to Recognize Fake News**: Eight felt confident in their ability to detect fake news.
- **Broader Themes Explored**
- **Access and Trust in AI Tools**: The discussion highlighted the need to make AI tools accessible to a broader audience, particularly seniors and less tech-savvy individuals. Building trust in these tools was recognized as critical, as even reputable fact-checkers often face skepticism.
- **Shades of Disinformation**: Participants acknowledged that false information exists on a spectrum, ranging from accidental misinformation to deliberate disinformation. The framing of messages, particularly through emotional language, influences how readers interpret content.
- **Role of Media and Collaboration**: Some advocated for collaboration between journalists and fact-checkers to enhance news accuracy, with an emphasis on transparency and accountability in the media.
- **Combatting Malicious Actors**: Participants discussed the challenges of identifying and countering deliberate disinformation campaigns, with insights shared from case studies on adaptive disinformation tactics, particularly those from Russia.
- **Education and Demystifying AI**: A call was made to promote public understanding of AI and its mechanisms, starting with early education. This could help foster curiosity and empower individuals to use AI tools effectively.

Funded by the European Union

The discussion underscored the complexity of addressing fake news and the need for a holistic approach involving technology, education, and collaboration across sectors. The event effectively fostered meaningful dialogue among stakeholders, addressing the project's objectives and exploring solutions to the challenges posed by disinformation.

## ○ FOCUS GROUP EVENT KIEV - REPORT.

On December 12, a focus group brought together key stakeholders actively engaged in combating disinformation both in Ukraine and internationally. Participants included journalists, analysts, media educators, and advisors from various Ukrainian cities. This diverse group shared their experiences, insights, and challenges in addressing the spread of disinformation and propaganda originating from the Russian Federation, particularly since 2022. Experts provided insights on topics such as disinformation trends, its connection to climate change, effective countermeasures, challenges faced, and recommendations for enhancing collaboration among stakeholders in the fight against disinformation.

Experts highlighted the proliferation of disinformation across different regions of Ukraine, with specific trends linked to audience demographics and platform usage. It was noted that disinformation sources vary by age group and social status: teenagers and elderly populations predominantly encounter it on TikTok, young adults (20-30 years) on Instagram and Telegram, and middle-aged and older adults (40+) on Telegram, Facebook, YouTube, as well as interpersonal networks like neighbors and colleagues. Despite these differences, platforms like Telegram, YouTube, and TikTok were identified as the primary sources of disinformation due to their widespread usage and algorithmic amplification.

Participants emphasized the effectiveness of prebunking, which involves proactively addressing false narratives before they gain traction, and debunking, which focuses on reactively clarifying misinformation with evidence-based corrections. Local media and educational initiatives focusing on media literacy were recognized as critical tools for building societal resilience against disinformation.

Several challenges were identified during the discussion. Misinformation often spreads faster and more widely than corrective content, undermining debunking efforts. Additionally, low trust in governmental communication and public institutions creates barriers to disseminating credible information. Analysts and fact-checkers frequently experience burnout due to the overwhelming volume of disinformation and the intensity of their work.

Participants underscored the importance of improving coordination among stakeholders, including government agencies, non-governmental organizations (NGOs), and international bodies. Proposed solutions included leveraging AI-driven analytics and maintaining white lists of credible media outlets to streamline fact-checking efforts, building public trust through transparent communication, and expanding educational projects to equip audiences with critical thinking skills and tools to identify disinformation.

The discussion concluded with a strong emphasis on fostering cross-sector collaboration and prioritizing trust-building initiatives. By integrating innovative technologies, enhancing media literacy, and strengthening partnerships between local and international organizations, stakeholders can more effectively counter the complex and evolving threat of disinformation.

Funded by the European Union

# Review Sheet of Deliverable/ Milestone Report

# D.12.5 : Multi-Stakeholders' perceptions

| | |
|---|---|
| **Editor(s):** | Gaborit Pascaline, Joen Martinsen, Vishnu Rao |
| **Responsible Partner:** | Pilot4Dev |
| **Status-Version:** | Final - v0.4 |
| **Date:** | 22/01/2025 |
| **Distribution level (CO, PU):** | Public |
| **Reviewer (Name/Organization)** | Arianna D'Ulizia (CNR), Alessia D'Andrea (CNR) |
| **Review date** | 28/01/2025 |

*Disclaimer: This assessment reflects only the author's views and the European Commission is not responsible for any use that may be made of the information contained therein"*

Funded by the European Union

Mark with X the corresponding column:

| Y= yes | N= no | N = not applicable |
|---|---|---|

| ELEMENT TO REVIEW | Y | N | NA | COMMENTS |
|---|---|---|---|---|
| **FORMAT**: Does the document … ? | | | | |
| …include editors, deliverable name, version number, dissemination level, date, and status? | X | | | |
| …contain a license (in case of public deliverables)? | X | | | |
| …include the names of contributors and reviewers? | X | | | |
| ….has a version table consistent with the document's revision? | X | | | |
| … contain an updated table of contents? | X | | | |
| … contain a list of figures consistent with the document's content? | | | X | The report does not contain figures. |
| … contain a list of tables consistent with the document's content? | | | X | The report does not contain tables. |
| … contain a list of terms and abbreviations? | X | | | |
| … contain an Executive Summary? | X | | | |
| … contain a Conclusions section? | X | | | |
| … contain a List of References (Bibliography) in the adequate format, if relevant? | X | | | |
| … use the fonts and sections defined in the official template? | X | | | |
| … use correct spelling and grammar? | X | | | |
| … conform to length guidelines (50 pages maximum (plus Executive Summary and annexes) | X | | | |
| … conform to guidelines regarding Annexes (inclusion of complementary information) | X | | | |
| … present consistency along the whole document in terms of English quality/style? (to avoid accidental usage of copy&paste text) | X | | | |
| **About the content…** | | | | |
| Is the deliverable content correctly written? | X | | | |
| Is the overall style of the deliverable correctly organized and presented in a logical order? | X | | | |

Funded by the European Union

| ELEMENT TO REVIEW | Y | N | NA | COMMENTS |
|---|---|---|---|---|
| Is the Executive Summary self-contained, following the guidelines and does it include the main conclusions of the document? | X | | | |
| Is the body of the deliverable (technique, methodology results, discussion) well enough explained? | X | | | |
| Are the contents of the document treated with the required depth? | X | | | |
| Does the document need additional sections to be considered complete? | | X | | |
| Are there any sections in the document that should be removed? | | X | | |
| Are all references in the document included in the references list? | X | | | |
| Have you noticed any text in the document not well referenced? (copy and paste of text/picture without including the reference in the reference list) | | X | | |
| **SOCIAL and TECHNICAL RESEARCH WPs (WP4, 5, 12, 13, 14)** | | | | |
| Is the deliverable sufficiently innovative? | X | | | |
| Does the document present technical soundness and its methods are correctly explained? | X | | | |
| What do you think is the strongest aspect of the deliverable? | | | | It provides a set of open challenges and recommendations useful for the other social and technological WPs.<br>It is very clear, interesting and consistent. |
| What do you think is the weakest aspect of the deliverable? | | | | None |
| Please perform a brief evaluation and/or validation of the results, if applicable. | | | X | |
| **AI AND TECNOLOGICAL WPS (WP6 – WP11 )** | | | | |
| Does the document present technical soundness and the methods are correctly explained? | | | | |
| What do you think is the strongest aspect of the deliverable? | | | | |

Funded by the European Union

| ELEMENT TO REVIEW | Y | N | NA | COMMENTS |
|---|---|---|---|---|
| What do you think is the weakest aspect of the deliverable? | | | | |
| Please perform a brief evaluation and/or validation of the results, if applicable. | | | | |
| **DISSEMINATION AND EXPLOITATION WPs (WP15 – WP17)** | | | | |
| Does the document present a consistent outreach and exploitation strategy? | | | | |
| Are the methods and means correctly explained? | | | | |
| What do you think is the strongest aspect of the deliverable? | | | | |
| What do you think is the weakest aspect of the deliverable? | | | | |
| Please perform a brief evaluation and/or validation of the results, if applicable. | | | | |

**SUGGESTED IMPROVEMENTS**

| PAGE | SECTION | SUGGESTED IMPROVEMENT |
|---|---|---|
| | | |

**CONCLUSION**

Mark with X the corresponding line.

| | |
|---|---|
| X | Document accepted, no changes required. |
| | Document accepted, changes required. |
| | Document not accepted, it must be reviewed after changes are implemented. |

Please rank this document globally on a scale of 1-5 (1 = poor, 5= excellent) – using a half point scale. Mark with X the corresponding grade.

| Document grade | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | X |

Funded by the European Union