FORUM INFORMATION DEMOCRACY

THE GOVERNMENT OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of State
Department of Media, Connectivity and Digital Policy

THE GOVERNMENT OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of Foreign and European Affairs, Defence, Development Cooperation and Foreign Trade

Ministry of Culture of Ukraine

Meeting of the Partnership for Information and Democracy

Workstream on Safeguarding Access to Reliable Information in the Age of AI

1st Meeting: April 29, 2026

Summary

First Meeting: Setting the scene

29 April - 2 to 3:30pm CEST

1/ Overview

On 29 April 2026, the Partnership for Information and Democracy convened its first closed-door session of the workstream on Safeguarding Access to Reliable Information in the Age of AI. Co-chaired by Ukraine and Luxembourg, the meeting provided insights into the existing state-of-the-art on AI in the information space. Representatives of signatory states and of civil society presented existing challenges and open research questions.

2/ Opening Remarks

- **Camille Grenier,** Executive Director at the Forum on Information and Democracy, welcomed participants and thanked colleagues, partner states, co-chairs and the rapporteur Can Şimşek for the preparation of the Background Report. He presented the two pillars of this new workstream: the first on the impact of AI on media in terms of sustainability and loss of traffic (up to 80%), and the mutual dependency between AI and media. The second on how AI content affects knowledge and trust in society. He announced that the workstream would continue in other arenas such as the Paris Peace Forum, Internet Governance Forum and the Yerevan Dialogue among others.

- **Luc Dockendorf,** Ambassador for Cybersecurity and Digitalisation of Luxembourg, seconded Mr. Grenier's remarks and thanked Mr. Şimşek for the well-written background report. He cited the Paris Declaration and reiterated Luxembourg's longstanding support extended to the FID. He cited several existing strategies on Digital Economy and AI topics and

stressed the importance of collaboration on this framework, both via this workstream and the 2025 workstream on Private Messaging applications (of which Luxembourg was also a co-chair). He reiterated that disinformation is a matter of life and death for many populations globally, especially in Ukraine in the context of Russian generative-AI propaganda in the current aggression, as well as other actors globally complicit in crimes against humanity and genocide. Hence, there is a need to discuss these topics in a multistakeholder setting. He also stressed the need to remember that the work is carried out in the interest of those most vulnerable and most affected by these unconstrained economies.

- **Ganna Krasnostup,** Director of the Department of Strategic Communications and Promotion of Ukrainian Culture at the Ministry of Culture of Ukraine, expressed her honour at Ukraine's co-chairpersonship. She stated that AI-augmented threats to the information space are not abstract and are a fundamental pillar of collective security. Ukraine faces the daily reality of both a physical and a digital war. This adds a complex layer to the ongoing challenges of Foreign Information Manipulation and Interference (FIMI). She cited as examples the systematic use by Russia of deepfakes, synthetic voices and images to manipulate public trust and strength, and the increasing difficulty in detecting such content. This risks ruining public trust in political systems and impacts the very foundations of democracy. She also stated that beyond the threats from disinformation, there is a need to understand how AI reshapes media landscapes. In Ukraine, media sources are a vital lifeline for citizens. Chatbots such as ChatGPT or Gemini drive audiences away from original news sources. She stated several key objectives for the year:
    - to define and address the legal aspect of information manipulation,
    - to move beyond reactive measures to proactive measures to ensure responsible AI,
    - and to ensure representation and pluralism are protected, and control by dominant players is prevented.

  She noted that the existing document isn't merely a policy document but a roadmap to ensure that AI supports and does not compromise access to quality information. She closed by saying that together, we must treat information as a public good.

3/ Setting the Scene and presenting the background report

Can Şimşek, consultant and rapporteur for the workstream, provided insights from the background report on the extant and emerging questions and challenges of AI in the information space.
The presentation delved into the central question of how to ensure access to reliable information via two pillars of policy research: one on the economic model of journalism and AI tools; and the other on the epistemic risks such as AI-driven manipulation.

- He laid down the key **economic issues** including;
    - Media visibility, viability and pluralism
    - Legal status of AI training materials, including the content subject to intellectual property law
    - Transparency
    - Fair compensation of journalists, data workers and related rights holders
    - Reputation, trust and autonomy of media publishers
    - Imbalance between authentication and generation costs including hyperrealistic content

- As for **potential policy interventions**, he mentioned several directions including:

    - Transparency and provenance
    - Fair remuneration
    - Stronger copyright regimes and licensing
    - Levy and digital service taxes

As for the second aspect on **epistemic risks and information integrity**, he underlined the differences between benign persuasiveness and manipulation as well as outcome harms vs. process harms. He underlined that certain capabilities and contexts merit particular attention, including:

- Hyperrealistic content, which enables fraud, manipulation and undermining of trust in verified information
- Agentic AI, that automates and scales information manipulation
- Sensitive contexts, such as disasters, elections, armed conflict, emergencies etc. where even short term events can have long term consequences

Finally, he underlined that the challenge is not merely to counter “fake news,” but to safeguard systematic access to reliable information and to protect public trust in democratic institutions and deliberative processes.
In terms of policy direction, he emphasized that there is no single legal remedy. Effective responses must combine a range of instruments, from risk based frameworks such as the EU Digital Services Act and AI Act, to privacy regulation and criminal law where appropriate. He further noted that technical measures, including labeling, watermarking, and provenance mechanisms, are essential. Lastly, he stressed the need to educate and inform citizens about the realities of

this new information ecosystem.

Mindaugas Stanys, Head of Strategic Communications Division at the Ministry of Foreign Affairs of Lithuania, confirmed the findings of the background report as corresponding to their own discussions with media organisations and investigations.

- He reiterated that AI is changing how information is accessed, discovered and trusted. He highlighted that AI is boosting existing information manipulation tactics and being an enabler of FIMI in Lithuania and Eastern Europe.

- He added that such FIMI operations seek to undermine not only trust in institutions and election processes but also leadership or even the existence of countries.

- In particular, he mentioned historical revisionism by the use of generative AI: The information environment is being flooded with false information which is then used by AI to answer user queries. He concluded that legal and other policy interventions safeguarding transparency and accountability are needed on national and international levels.

4/ Presentation of key elements from the AI as a public good report

Katharina Zuegel, Policy Director at the Forum on Information and Democracy, presented the organization's 2024 report on AI as a public good, affirming the continued relevance of its recommendations. She outlined that AI systems must be safe, incorporating risk mitigation measures, as well as moderated, fact-checked, ethical, accountable through clear governance rules, transparent, governed through participatory processes, and respectful of data governance rules, intellectual property rights, and privacy.

- Zuegel stressed that media organizations must be informed when their content is used for AI training or retrieval, and must be granted both the right to opt out and the right to fair compensation. On liability, she argued that AI developers and deployers should be held accountable when they fail to comply with applicable rules, and that platforms hosting AI systems should face equivalent liability regimes.

- She raised the question of responsibility for AI-generated content, noting that unlike third-party content, such material is produced by the AI systems themselves. She also underlined the importance of users' rights to be informed, to receive explanations for AI-driven decisions, and to be protected from discrimination.

- Zuegel noted that regulation alone is insufficient. Governments can leverage public procurement to promote ethical standards in AI development and use. She also called for certification systems analogous to fair trade labels, enabling users and institutions to identify ethical AI

products, alongside broad investment in AI and digital literacy.

- On governance, she discussed the establishment of a **nodal authority**, whether an existing body with an expanded mandate or a newly created institution, tasked with overseeing AI regulation. She advocated for strict, tiered transparency rules, legal pathways enabling independent researchers to audit AI models, and accessible mechanisms for legal redress. Finally, she proposed financing these oversight structures through a tax on AI company revenues, applying a polluter-pays principle.

5/ Case studies

5.1/ Media sustainability & viability

Dorien Verckist, Senior Media Analyst and Public Value Lead at EBU-MIS, presented findings from the collaborative BBC and EBU research initiative News Integrity in AI assistants, which examined the news quality in AI assistant responses.

- Building on an earlier BBC study from February 2025, the project expanded the analysis internationally, involving 22 public service media organisations across 18 countries and 14 languages. The study assessed responses from ChatGPT, Perplexity, Gemini, and Copilot, using their free versions to reflect the baseline experience available to ordinary users.

- The findings showed that AI assistants still perform unreliably as news intermediaries. **45% of responses contained at least one significant issue**, including factual inaccuracies, missing context, weak or absent sourcing, misleading attribution, and cases where opinion was presented as fact. Around **20% of responses had significant accuracy problems**, while **31% had serious sourcing issues**, making it difficult for users to verify information or identify the original news source.

- Verckist stressed that these problems are not limited to one country or broadcaster, but represent a **wider structural risk to news integrity across languages and markets**. As younger audiences increasingly use AI assistants for news, unreliable summaries may undermine public trust, damage the reputation of trusted broadcasters, and weaken the visibility of plural, authoritative journalism.

- The presentation also introduced the accompanying EBU toolkit, which maps recurring AI failure modes and proposes standards for better practice. The broader “Facts In, Facts Out” campaign calls for cooperation between media organisations and technology companies around three core principles: authorisation, attribution, and plurality of news sources.

5.2/ Gen AI for Information Manipulation

Tetiana Avdieieva, Senior Legal Counsel at Digital Security Lab Ukraine (DSLU) and an Expert at the Expert Committee on AI under the Ministry for Digital

Transformation of Ukraine, presented her findings on AI use in wartime media and freedom of expression. She stressed that AI use in and around the media has dual implications for the information environment, especially from a FIMI perspective.

- On the positive side, AI can strengthen media resilience. It can support pluralism, fact checking, large scale data analysis, content moderation, transcription, translation, and newsroom efficiency. In Ukraine, where Russia's aggression has led many journalists and media workers to be drafted or displaced, AI tools may help under-staffed local and smaller media outlets continue operating.

- However, Avdieieva warned that AI is also being used maliciously. Beyond the more visible risks of deepfakes, voice cloning, and synthetic content, AI can enable the coordination, scaling, and targeting of manipulation campaigns. It is also increasingly used in attacks on information infrastructure, including phishing, device compromise, remote access operations, and attacks against journalists, civil society actors, government officials, and political leaders. She argued that these infrastructural threats are under addressed because public and regulatory attention tends to focus on visible synthetic content. Measures such as content marking, fact checking tools, and transparency obligations may be more enforceable for good faith media actors, but they do not sufficiently address covert, targeted, and technically sophisticated attacks.

- She highlighted two examples. **ClickFix** uses social engineering to trick users into approving malicious commands, often through fake software updates or bug fix prompts, resulting in malware installation. Secondly, **"Portal Kombat"** involves networks of malicious websites impersonating trusted media outlets such as the BBC, The New York Times, The Guardian, Ukrainian media, and local European outlets. **These sites may copy real articles or publish forged and generated content, especially around sensitive topics.** Because AI assistants increasingly retrieve information from online searches, such websites can pollute AI generated news summaries and make propaganda appear credible.

- She also noted examples from social media, where Russian actors use images of Ukrainian television producers or media brands to build audiences for apparently legitimate channels, before using those channels to distribute hostile or manipulative narratives.

- Finally, she emphasized the need for stronger cooperation between media, civil society, governments, and technology platforms. This includes better access to platform communication channels, faster escalation procedures, technical and legal expertise, and context specific knowledge. She argued that Ukraine's non EU status limits some available regulatory tools and this creates an important space for deeper collaboration.

5.3/ AI and Elections in Brazil

Liz Nóbrega, Coordinator of Strategic Communications and Innovation at Alafia Lab, presented findings from the Observatório de IA nas Eleições, which

monitored the use of generative AI during Brazil municipal elections, held in October 2024.

The Observatory mapped AI uses between 16 August and 31 October 2024, monitoring platforms, press outlets, and fact checking organisations. It found that AI was used both by official campaigns and by voters, across images, videos, audio, jingles, narration, chatbots, and synthetic campaign material. The latest iteration of this report, published in parts since, analyses cases identified between December 2025 and February 2026, nine months before the Brazilian elections. She highlighted several areas of concern from both iterations:

- Disinformation and deepfakes, including voter generated synthetic videos targeting candidates. The second iteration saw **50% more incidents compared to the average identified during the previous year**. Most of this content circulated without any transparency. **Only 27% of the identified materials included labels, watermarks, or any indication that AI had been used.** The remaining synthetic political content circulated online without any warning to users, including posts shared directly by politicians and political parties themselves.

- **Chatbot systems rank political candidates, recommend candidates to users, and make qualitative judgments about political proposals**. This raises important concerns because these systems increasingly mediate access to political information and may influence electoral perceptions in subtle but powerful ways.

- Approximately **one-third of the identified publications originated from political actors**. This included posts from elected officials, political influencers, and party accounts. **45% of the identified cases had clearly disinformative characteristics, while 55% circulated mainly in contexts of satire or humor.** However, even humorous content can shape political perceptions, normalize hostility, and reinforce political attacks. **More than 60% of the documented cases were used to criticize or attack political** opponents.
- Audio deepfakes, which are especially difficult to verify quickly and accurately, particularly when circulated through messaging apps.

- Deepnudes targeting female candidates; among the most concerning examples were manipulated images targeting congresswoman Érika Hilton, a deepfake portraying former First Lady Michelle Bolsonaro as a prostitute, and AI-generated erotic profiles used to support political candidates. These cases demonstrate how generative AI is increasingly being weaponized through gendered and sexualized political violence.

- Chatbots and AI assistants, including generative AI platforms such as Google Gemini and Meta AI providing incorrect or outdated electoral information.

- Uneven enforcement

Looking forward, Nóbrega **highlighted the emergence of fully AI generated political actors** such as “Dona Maria”, an influencer avatar representing a black

working-class woman. She underlined that the people who are engaging with this account are not necessarily aware of the fact that it is synthetic content. She also underlined that **Brazil's Superior Electoral Court introduced specific AI rules for the elections**, including requirements for labelling AI generated content, restrictions on avatars and chatbots, and prohibitions on deepfakes. She noted that they are expanding this research to AI-enabled attacks on human rights.

5.4/ Information Integrity in Asia

Syed Nazakat, Founder and CEO of DataLEADS, spoke about the growing threat of deepfakes and information integrity challenges in Asia, with particular attention to elections, finance, and public trust.

- He argued that India is facing a rapidly intensifying deepfake problem. DataLEADS has been tracking AI enabled disinformation across electoral politics, banking, financial institutions, and other sectors where synthetic media can cause public harm.

- In the 2024 Indian election context, DataLEADS noted that AI generated content remained a relatively small share of overall misinformation (around 3%), but included deepfakes of deceased politicians, false celebrity endorsements, and synthetic material involving Bollywood figures.

- DataLEADS has also been examining deepfake and synthetic media risks across Bangladesh, India, Indonesia, and the Philippines. Across these South and Southeast Asian contexts, he suggested that the principal targets are similar: politics, business, financial systems, and influential public figures.

- The key change, compared with only two years ago, is that **misinformation is no longer primarily produced through conventional image, video, or textual manipulation. AI generated and AI manipulated content is becoming a qualitatively different threat.**

- He emphasised that deepfakes have real electoral consequences. One example concerned **Bangladesh**, where a synthetic video reportedly circulated in 2024 shortly before national elections voting and falsely suggested that a candidate was withdrawing from the election. Such cases matter because voters may encounter manipulated content at moments when there is little time for verification, correction, or institutional response.

- Nazakat noted that deepfake risks are not confined to politics. Politically motivated deepfakes are growing, but financial misuse is already a major vector. The *Contours of Cybercrime 2025* report, published by DataLEADS, revealed that Indians lost an estimated $2.66 billion to cyber fraud in 2024, with a significant share of these crimes driven by emerging AI-enabled threats such as deepfakes, voice cloning, and AI-powered phishing campaigns. Nearly one-third of the misinformation content analysed contained AI-generated media — including manipulated videos, synthetic audio, and fabricated images — highlighting the rapidly evolving sophistication of digital deception and the growing challenges posed by

generative AI to information integrity and public trust.

- He further observed that platform dynamics vary by country. In some contexts, deepfake content is more mature and visible on Facebook, while in Indonesia, TikTok appears especially important. This reflects broader differences in platform use, language communities, and electoral communication practices across Asia.

- He also warned about **foreign linked influence operations in Asia**. He referred to China and Russia linked campaigns, including multilingual state linked networks designed to undermine trust in democratic processes and public institutions. He connected these concerns to wider geopolitical information operations, including narratives around Taiwan and the Iran war. He noted that work on these campaigns was ongoing, with further findings expected later.

- A major challenge, in his view, is **linguistic and regional asymmetry**. Many AI tools, detection systems, and platform moderation systems are built in Western contexts and do not adequately understand Asian languages, dialects, political references, or cultural cues. This creates serious gaps in detection, moderation, and response capacity.

- To address this, DataLEADS is working with Indian technical institutions, including IIT Jodhpur and IIT Madras, on **deepfake detection capacity**, with support from India's Ministry of Electronics and Information Technology(MeitY). Nazakat described a planned detection tool trained on a large dataset and designed to cover Indian languages. He also mentioned **Verify.AI**, a forthcoming tool by DataLEADS for real-time social listening, multilingual intelligence, and collaborative verification.

- His broader conclusion was that **Asia is experiencing a sharp rise in deepfake and synthetic media threats, but the region lacks sufficient resources, language specific tools, institutional capacity, and rapid response mechanisms.**

6/ EU and US Copyright Regime

6.1/ Anya Schiffrin's intervention was centred on the OPD working paper *How to Update EU and US Copyright Regimes in the Age of AI*, co-authored with Roberta Carlini and Natalia Menéndez. The paper examines how existing copyright rules in the United States and the European Union are being stretched by generative AI. In particular, it looks at the use of copyrighted material for training data, text and data mining, and retrieval augmented generation (RAG), as well as the emerging market for licensing agreements between publishers and AI companies. She also referred to her related chapter with Haaris Mateen, *How to Calculate What News is Worth to AI*

- Schiffrin focused especially on the US legal context. She noted that the situation remains shaped by the broad doctrine of fair use, which gives AI companies significant room to argue that using copyrighted material for

training is lawful. However, **the field is still unsettled, with cases such as *The New York Times v OpenAI and Microsoft* likely to be especially consequential.**

- As for the EU, **Roberta Carlini**'s part of the work turned to the Copyright in the Digital Single Market Directive, especially the text and data mining exceptions in Articles 3 and 4, and asked whether these rules are adequate for AI training and AI generated outputs.

- She then discussed the emerging European case law. In *Kneschke v LAION*, decided in Hamburg in 2024, the court accepted the application of the TDM exception. In *GEMA v OpenAI*, before the Munich court, the litigation raises different questions about copyrighted works, generative AI, and infringing outputs. She also referred to the Dutch dispute involving DPG and HowardsHome, and to forthcoming clarification from the CJEU, especially in *Like Company v Google Ireland*, which may help determine how TDM applies in platform and search contexts.

- Schiffrin noted that this is also a competition and market power problem. From December 2025, the European Commission began examining possible anti competitive conduct linked to the use of content by AI systems. A similar concern has appeared in Brazil, where the competition authority has also investigated how dominant digital platforms use journalistic content.

- She further stressed that the contrast between jurisdictions is significant. In the United States, there is extensive litigation, with AI companies relying heavily on fair use. In Europe, by contrast, a licensing market is beginning to develop, with agreements between some publishers and generative AI companies. This market may have a positive effect on media financing, but it may also harm media pluralism. The agreements appear to be concentrated in large countries and large media markets, and their terms are generally confidential. We do not know who is being paid, how much, for what content, or under what conditions. If only large publishers can negotiate with AI firms, the emerging market may strengthen already powerful actors while leaving smaller, local, or minority language media outside the system.

- Schiffrin also reminded current European policy initiatives. She referred to reports by Lucchi, Peukert, Voss, and Jensen, all of which address the need to change the legal regime, close the value gap, and secure remuneration for right holders. One proposed solution is a form of statutory or compulsory licensing. The European Parliament's vote of 10 March 2026 was presented as part of this trend, since it supported the creation of a functioning market for AI uses of copyrighted works. She noted that similar concerns also appear in Council of Europe discussions. However, she questions whether a real market could be created when there is such an imbalance of power between AI companies and rights holders?

- She concluded that, although many reform proposals now exist internationally, the US remains more constrained. Statutory relief may be politically difficult there, even though compulsory licensing has worked in

other sectors. The harder question is whether such models can be adapted to generative AI, given the number of intermediaries and the opacity of value extraction. As she noted, the problem is not conceptually insoluble; it is a matter of designing a workable system and securing agreement among the relevant actors.

6.2/ Suzanne Vergnolle, Associate Professor CNAM, structured her intervention around three common misconceptions about the EU AI Act.

- First, she challenged the idea that the EU regulates AI mainly through the AI Act. The AI Act is a cross sectoral, risk based regulation, but it does not exhaust the existing EU legal framework for AI. AI systems are also governed by other instruments, including the GDPR, especially on automated decision making, the Digital Services Act, and relevant national laws. The AI Act therefore sits within a broader regulatory ecosystem. Its core logic is risk classification: prohibited practices, high risk systems with numerous obligations, limited risk systems subject mainly to transparency duties, and minimal risk or no risk systems with no specific AI Act obligations.

- Second, she warned that reading the AI Act alone is not enough to prepare for and understand compliance. Many obligations depend on annexes, guidelines, standards, and implementing documents. For example, Article 6(2) must be read together with Annex III, which lists high risk AI use cases. Similarly, the Commission has issued guidance on prohibited practices, while further documentation on general purpose AI, copyright related transparency, and compliance is part of the Act's implementation architecture.

- Third, Vergnolle rejected the assumption that the AI Act will operate identically across all Member States. Although it is an EU regulation, implementation will depend on national supervisory structures. Member States were required to designate competent authorities by 2 August 2025, but many have still to comply and governance models differ considerably. Some countries rely on fragmented sectoral regulators, while others move toward more centralised oversight. This makes compliance practically complex, especially where several authorities may share responsibility.

6.3/ Luc Dockendorf, took the floor and argued that regulation "on paper" will not stop criminals, dominant technology firms, or authoritarian actors. The strongest drivers of harmful AI use remain illicit financial gain, strategic power, and a pervasive narrative that AI development is inevitable. He urged civil servants and civil society actors to resist this inevitability narrative and to insist that AI deployment remains a matter of political choice, public accountability, and democratic contestation.

7/ Closing and next steps

Emma Gruden, Policy Officer at the Forum on Information and Democracy, concluded the meeting by welcoming the strong basis established for future work, including three forthcoming technical meetings. The next will take place on

18 June, with civil society coalitions invited to share their work and policy recommendations, and states invited to designate points of contact.